Mixing in Time and Space for Discrete Spin Systems

by

Dror Weitz

B.S. (Hebrew University of Jerusalem) 1999

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Computer Science

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Alistair Sinclair, Chair
Professor Richard Karp
Professor Yuval Peres

Spring 2004

Mixing in Time and Space for Discrete Spin Systems

Abstract

Mixing in Time and Space for Discrete Spin Systems

by

Dror Weitz

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Alistair Sinclair, Chair

This dissertation studies relationships between fast convergence to equilibrium (mixing in time) of natural Markov chain Monte Carlo algorithms for discrete spin systems, and decay of correlations with distance in the corresponding equilibrium distribution (mixing in space). The results fall into four main groups.

In the first part we generalize the *Dobrushin* and *Dobrushin-Shlosman* conditions for uniqueness of the Gibbs measure (a form of mixing in space) by presenting conditions of any finite size for models on any underlying graph. We give two dual conditions, one requiring that the total influence *on* a site is small, and the other that the total influence *of* a site is small. As for the original ones, our conditions also imply $O(n \log n)$ mixing time of the corresponding Markov chain, and this connection is discussed in detail. In contrast to the proofs of the original conditions, our proofs are combinatorial in nature and use tools from the analysis of discrete Markov chains, in particular the path coupling method.

In the second part we critically examine a known sharp equivalence between appropriate notions of mixing in time and in space. For this part, the discussion applies only to systems on the d-dimensional integer lattice $\mathbb{Z}^d$. We give new, purely combinatorial arguments to prove that, if the mixing time of the Glauber dynamics is $O(n \log n)$, then spin correlations decay exponentially fast with distance in the Gibbs distribution. We also prove the converse implication for monotone systems, and for general systems we prove that exponential decay of correlations implies $O(n \log n)$ mixing time of a dynamics that updates sufficiently large blocks (rather than single sites). While the above equivalence was already known to hold in various forms, our proofs avoid the functional analysis machinery

employed in previous proofs.

In the third part we develop a new framework for analyzing the mixing time for spin systems on trees. The main technical result here is that on trees, an appropriate form of mixing in space implies $O(n \log n)$ mixing time of the Glauber dynamics. The novelty of this implication is that it is specific to the boundary condition. This allows us to give the first comprehensive analysis (in any context) of the effect of boundary conditions on the mixing time for the Ising and other models. Specifically, for the Ising model we show that the mixing time on an n-vertex regular tree with $(+)$-boundary remains $O(n \log n)$ at all temperatures (in contrast to the free boundary case, where the mixing time is not bounded by any fixed polynomial at low temperatures). We also show that this bound continues to hold in the presence of an arbitrary external field. Our results are actually stronger, and provide tight bounds on the log-Sobolev constant and the spectral gap of the dynamics. In addition, our methods yield simpler proofs and stronger results for the mixing time in the regime where it is insensitive to the boundary condition. We apply our techniques to other models as well, and obtain $O(n \log n)$ mixing time over a significantly wider range of parameter values than previously known for independent sets, colorings and the Potts model. This includes situations in which the mixing time is strongly dependent on the boundary condition, as well as situations in which fast mixing is proved for all boundary conditions.

In the fourth part we explore directions for extending our results for trees to the 2-dimensional integer lattice $\mathbb{Z}^2$. The main motivation here is resolving a long standing conjecture which states that, conditioned on the all-$(+)$ boundary, the mixing time remains bounded by a fixed polynomial in n at all temperatures. (Notice that for the free boundary case, the mixing time at low temperatures is known to be very slow, specifically $\exp(\Theta(\sqrt{n}))$.) We present a new implication for systems on $\mathbb{Z}^2$, where a certain form of mixing in space implies polynomial mixing time of the dynamics, and the implication is specific to the boundary condition. Although the form of mixing in space in the hypothesis of our condition is still too strong to have an immediate application, it suggests directions for future research towards resolving the above conjecture.

To my mother

Contents

List of Figures

Acknowledgments

First and foremost, I would like to thank my advisor, Alistair Sinclair for his guidance, support, and attention to detail. His approach and the intriguing discussions we had on both mathematics and other topics have greatly influenced the way I conduct research, present my ideas, and view life in general.

I am greatly indebted to Fabio Martinelli, co-author and friend. It was during his year at Berkeley that I flourished; the ping-pong of ideas that we engaged in exposed me to the virtues of conducting research in an interdisciplinary field, and I learned greatly from his knowledge. I would like to thank my co-authors Eric Vigoda and Martin Dyer for sharing with me their earlier work on mixing. I also thank Elchanan Mossel for being there whenever I needed to share some ideas or discuss the progress I made when no-one else understood what I was talking about.

I would like to recognize the members of my dissertation committee, Yuval Peres and Richard Karp, for their useful advice and support. I would also like to acknowledge Luca Trevisan for his role in my qualification exam, and Mary Byrnes for working her magic to ensure that the exam took place.

I am grateful to the faculty members of the Theory Group for making the group so vibrant and diverse. I especially want to thank Satish Rao for numerous meaning-of-life conversations and for showing me what Christmas spirit is all about.

I was very fortunate to be part of an amazing group of students (and postdocs) at the Theory Group. I would like to thank all my friends who hung around the theory offices (and with whom I often spent time off campus as well), especially Jittat Fakcharoenphol, Eran Halperin, Samantha Riesenfeld, Iordanis Kerenidis, Ranjit Jhala, Ziv Bar-Yossef, Steve Chien, Kunal Talwar, Kevin Chen and David Ratajczak. Each of you inspired me in different ways, even if it sometimes took me a while to realize it.

Pursuing a doctorate degree and conducting this research would not have been possible without financial support. For this I thank the National Science Foundation[1], the University of California for a Regents' Fellowship, and the Theory Group at Microsoft Research where I spent a very productive summer.

An integral part of my graduate studies was the Berkeley experience, one that was shaped by the many people I had the privilege to get acquainted with throughout my time

[1]This work was supported in part by grants CCR-9820951 and CCR-0121555.

here. One of the most important experiences I had in Berkeley was living at the *International House*, to which I am grateful for providing me with a home away from home. At the I-House I had the opportunity to have friends from all over the world, of which I especially thank Merav, Erkan and Yuma. You were with me through moments of both joy and stress, and I will cherish your friendship forever. I thank my housemates in 2230 Virginia: Michal, Elchanan, Serban and Julia, and at the Shafter house: Biyi, Sean, Sarah, Corey, Danielle, Annelies and Dipti. All of you were like a second family for me, and sharing a house with you exposed me to numerous new aspects of life and also taught me a lot about myself.

I would like to recognize the various coffee shops in which this thesis was written, in particular *Brewed Awakening* and the *International House Café*, for supplying the power and the atmosphere that kept me writing. If I have broken the record for time spent on a single Mocha, I promise to compensate for this in the future when I come back as a tourist.

I owe special thanks to my long time friends from home, Guy, Ori and Udi. I was lucky to have you living in the US for the past few years (though on the wrong coast) so that we could share impressions of this far-away country. The legendary winter breaks we spent together pulled me out of a few rough spots; thanks for listening to the same complaints and dilemmas time and time again.

Finally, I thank my family, especially my parents, for their support over the years and for laying the solid foundation on which I rely wherever I go.

Chapter 1

Introduction

Spin systems are a class of models that originated in Statistical Physics, though interest in them has since expanded to many other areas, including Probability Theory, Statistics, Artificial Intelligence, Communication, and Theoretical Computer Science. A *spin system* consists of a collection of sites which are the vertices of an underlying connected graph. A *configuration* of the spin system is an assignment of one of a finite set of *spins* to each site. The sites interact locally, according to potentials specified by the system, such that different combinations of spins on neighboring sites have different relative likelihoods. This interaction gives rise to a well-defined probability distribution over configurations of any finite subset (region) of the sites. Such a distribution is referred to as a *finite volume Gibbs distribution*, and is regarded as the equilibrium state of the given subset.

For example, in the Ising model (probably the most widely considered spin system) on a finite graph $G = (V, E)$, a configuration $\sigma = (\sigma_x)$ consists of an assignment of ± 1-values to each vertex of V. The probability of finding the system in configuration $\sigma \in \{\pm 1\}^V \equiv \Omega_G$ is given by the *Gibbs distribution*

$$\mu_G(\sigma) \propto \exp\Big(\beta \sum\nolimits_{xy \in E} \sigma_x \sigma_y\Big), \tag{1.1}$$

where $\beta \geq 0$ is the inverse temperature. Thus μ_G assigns higher probability to configurations in which many neighboring spins are aligned. This effect increases with β, so that at high temperatures (low β) the spins behave almost independently, while at low temperatures (high β) there is global order. Frequently one imposes a *boundary condition* on the model, which corresponds to fixing the spin values at some specified "boundary" vertices of G; the term *free boundary* is used to indicate that no boundary condition is specified.

In the classical Ising model, $G = G_n$ is a cube of side $n^{1/d}$ in the d-dimensional Cartesian lattice $\mathbb{Z}^d$, and one studies the properties of the Gibbs distribution as $n \to \infty$ with a specified boundary condition (e.g., the all-$(+)$ or the all-$(-)$ configuration) on the faces of the cube; this limit is referred to as the "(infinite volume) Gibbs measure" for the given boundary condition. It is well known that a *phase transition* occurs at a certain critical inverse temperature $\beta = \beta_c$ (which depends on the dimension d): for $\beta < \beta_c$ (the "high temperature" region) there are no long-range correlations between spins and consequently there is a *unique* Gibbs measure independent of the boundary condition, while for $\beta > \beta_c$ (the "low temperature" region) correlations are present at arbitrary distances and there are (at least) two distinct Gibbs measures (or "phases"), corresponding to the $(+)$ and $(-)$-boundary conditions respectively. See, e.g., [Geo88, Sim93] for more background.

While the classical theory focused on static properties of the Gibbs measure, in modern statistical physics the emphasis has shifted towards dynamical questions with a computational flavor. The key object here is the *Glauber dynamics*, a Markov chain on the set of spin configurations Ω_G of a finite graph G. For definiteness, we describe the "heat-bath" version of Glauber dynamics: at each step, pick a vertex x of G u.a.r., and replace the spin at x by a random spin drawn from the distribution of σ_x *conditional on* all the neighboring spins. It is easy to check that the Glauber dynamics is an ergodic, reversible Markov chain on Ω_G whose stationary distribution is exactly μ_G. The Glauber dynamics is much studied for two reasons: firstly, it is the basis of Markov chain Monte Carlo algorithms, widely used in computational physics for sampling from the Gibbs distribution; and secondly, it is a plausible model for the actual evolution of the underlying physical system towards equilibrium. In both contexts, the central question is to determine the *mixing time*, i.e., the number of steps until the dynamics is close to its stationary distribution.

Advances in statistical physics over the past decade have led to the following remarkable characterization of the mixing time on finite n-vertex cubes with free boundary in the 2-dimensional lattice $\mathbb{Z}^2$ [SZ92, MOS94, MO94a, Mar98, Ces01, CGMS96]: when $\beta < \beta_c$ the mixing time is $O(n \log n)$, while for $\beta > \beta_c$ it is $\exp(\Omega(\sqrt{n}))$. Thus the phase transition (a static, spatial phenomenon) has a dramatic computational manifestation in the form of an explosion from optimal to exponential in the running time of a natural algorithm. This result stands as perhaps the most convincing example to date of an intimate connection between phase transitions and computational complexity, a connection that has recently received a lot of attention, in particular for k-SAT and related problems

(see, e.g., [FB99, BMZ02, AP03]).

The strikingly sharp correspondence in the above result between notions of temporal and spatial mixing is the principle motivation for this dissertation. By *temporal mixing* we mean that the Glauber dynamics converges "very fast" to its stationary Gibbs distribution, while by *spatial mixing* we mean that in the Gibbs distribution, correlations between the spins of different sites decay "very fast" with the (graph) distance between them. Beside suggesting a connection between phase transitions and computational complexity, this space-time correspondence is a rigorous example of another common heuristic observation in computer science: a "local algorithm" works well if and only if the problem it solves is of a "local nature". Here the local algorithm is the Glauber dynamics, and the problem it solves (sampling from the Gibbs distribution) is of a local nature depending on whether distant spins are correlated or not.

Correspondences between temporal and spatial mixing have been studied before. Most notable is the known equivalence between appropriate notions of temporal and spatial mixing for systems on the integer lattice $\mathbb{Z}^d$ [SZ92, MO94a, MO94b, Ces01]. (In fact, this equivalence was one of the main ingredients in achieving the sharp characterization of the mixing time for the Ising model on the square lattice described above.) An appropriate notion of temporal mixing is also known to imply a certain decay of correlations in general graphs [KMP01]. The motivation for further studying temporal-spatial correspondences in this dissertation stems from a need for progress on three levels. First, we seek a better understanding of the mathematical principles underlying this intriguing phenomenon. Second, we are interested in extending and generalizing the known correspondences between the two types of mixing, where the ultimate goal is to determine the full scope of this equivalence. Finally, establishing new correspondences between the two types of mixing is motivated by applications, where one type of mixing is known to hold but the other is still open. We present results that address all three of the above motivations: we provide new (simpler) proofs for known results, extend and generalize existing tools for establishing both types of mixing, and present new correspondences between spatial and temporal mixing using which we prove new rapid mixing results in various interesting scenarios.

Before we go on to discuss our results, it is worth mentioning that proving rapid mixing of the dynamics using spatial mixing properties of the stationary distribution is an example of a more general approach of using properties of the stationary distribution in order to analyze the dynamics (in contrast to elementary probabilistic techniques such as

coupling, which analyze the progress two arbitrary chains make towards coalescence in a single step of the chain). The idea here is that an analysis that uses properties of the stationary distribution is less likely to be affected by "locally bad configurations", which are detrimental to an analysis that considers the progress made in one step of arbitrary instances of the chain. Other examples of this approach include bounds on the mixing time stemming from the *conductance* of the Markov chain [SJ89] or from the congestion of a flow along its paths [Sin92], as well as a series of papers [DF01, Mol02, Hay03, DFPV04] that analyze the Glauber dynamics for sampling proper colorings, where the argument relies on the fact that after a sufficient "burn-in" time, the chain is very unlikely to be in a "bad" configuration.

Dobrushin type conditions

The first results we present are generalizations of the Dobrushin [Dob70] and Dobrushin-Shlosman [DS85a] conditions for uniqueness of the Gibbs measure. As their name suggests, conditions of this type were originally introduced as tools for establishing uniqueness of the infinite volume Gibbs measure (a form of spatial mixing), but are also known to imply rapid mixing of the dynamics. The fact that the same condition implies both spatial and temporal mixing (by independent arguments) is what makes us particularly interested in this tool.

The Dobrushin condition has been widely used to prove uniqueness of the Gibbs measure (and rapid mixing of the dynamics) in various models. What makes the condition so appealing is its simplicity, where only the single-site distributions of the system are considered. Thus, using a simple calculation based on a finite distribution, one is able to deduce properties of the infinite volume Gibbs measure (or the asymptotics of the mixing time of the dynamics in arbitrarily large regions). Though this condition does not capture all the scenarios in which the Gibbs measure is unique, achieving uniqueness bounds that extend beyond the ones established by this condition usually requires techniques that are far more complex. Dobrushin and Shlosman [DS85a] generalized the single-site condition by considering conditions which may depend on larger regions (but still of finite size). However, unlike the original Dobrushin condition, their condition is applicable only when the underlying graph of sites is the integer lattice $\mathbb{Z}^d$. Additional versions of the Dobrushin-Shlosman condition were given by others (e.g., Stroock and Zegarlinski [SZ92]), but still only in the context of $\mathbb{Z}^d$.

We generalize the above conditions by considering *both* larger volumes *and* any underlying graph. Naturally, all such conditions require that the influence spins at different sites have on each other is "small" in an appropriate sense. However, although they do not mention this explicitly, some of the conditions in the literature require that the total influence *on* a site is small, while others require that the total influence *of* a site is small. We make a clear distinction between these two cases, giving two dual conditions, both of them in the generality described above. In contrast to the proofs of the previous conditions, our proofs are purely combinatorial and are based solely on the technique of coupling two probability distributions. The arguments we use are of a dynamical nature, and thus our proof that spatial mixing follows from the conditions is very similar to that of deducing temporal mixing from them. We apply our conditions to prove uniqueness of the Gibbs measure for various models. Although the models we discuss are already known by other methods to admit a unique Gibbs measure, for most of them our results extend the range of parameters for which uniqueness is established using "finite size" conditions of the Dobrushin type.

Temporal and spatial mixing on the integer lattice

Although the Dobrushin-type conditions discussed above imply both spatial and temporal mixing, they do not establish a direct relationship between the two types of mixing (because the conditions are not necessarily tight). However, as mentioned above, a direct equivalence between the two types of mixing is known for systems on the integer lattice $\mathbb{Z}^d$. Specifically, it is known that $O(n \log n)$ mixing time of the Glauber dynamics (uniformly in the boundary condition) is equivalent to an exponential decay (with distance) of the influence of the boundary condition on the equilibrium distribution. The first results in this direction were given by Holley [Hol85] and Aizenman and Holley [AH87], followed by Zegarlinski [Zeg90] and culminating in the work of Stroock and Zegarlinski [SZ92], who were the first to establish the above equivalence in full. We further mention Martinelli and Olivieri [MO94a, MO94b], who later obtained sharper results by working with a weaker spatial mixing assumption, and Cesi [Ces01], who recently simplified some of the proofs. See also [Mar98] for a review of related results.

We shed new light on this equivalence by presenting new, simpler proofs of it. Furthermore, we explicitly point out the properties of the integer lattice used in the proofs (the essential property is that the volume of balls around a given point in $\mathbb{Z}^d$ is polynomial

in the radius rather than exponential), and isolate the arguments which use this property from those that hold in general graphs. The existing proofs in the references mentioned above use functional analysis, and usually discuss quantities such as the *spectral gap* and the *logarithmic Sobolev constant* of the dynamics as a measure of its temporal mixing (these quantities measure the contraction of the semi-group associated with the dynamics). Our proofs are purely combinatorial, based on the elementary technique of coupling probability distributions. Although some of the ideas we use have appeared before, ours is the first proof in which the whole argument is purely combinatorial. We note that the result we present in the direction going from spatial mixing to temporal mixing (of the single-site Glauber dynamics) is limited in the sense that it only applies to *monotone* systems. For general systems, however, we show that spatial mixing implies temporal mixing of a "finite-block" dynamics, in which a sufficiently large block of spins is updated at each step. The corresponding implication for the single-site dynamics in the general case is known [Ces01, Mar98, MO94b, SZ92], but currently we do not have a combinatorial proof of it.

Boundary-specific mixing

The notions of temporal and spatial mixing implied by the Dobrushin-type conditions discussed above, as well as the notions that are known to be equivalent for systems on $\mathbb{Z}^d$, are notions of mixing that are *uniform* in the boundary condition. Namely, these notions assert that the mixing time of the dynamics is $O(n \log n)$ under *any* boundary condition, or that correlations decay with distance in the Gibbs distribution conditioned on *any* boundary condition. One of the most interesting questions left open by the above results is establishing correspondences between temporal and spatial mixing that are sensitive to the boundary condition, and in particular, understanding the *influence of the boundary condition* on the mixing time. To better appreciate this point, let us go back to the Ising model on $\mathbb{Z}^2$. The Dobrushin-type conditions and the known equivalence mentioned above are relevant to the high temperature regime ($\beta < \beta_c$), where the mixing time is $O(n \log n)$ and correlations decay exponentially with distance uniformly in the boundary condition. Let us now shift our attention to the low temperature regime ($\beta > \beta_c$), where under the free boundary condition the mixing time is $\exp(\Omega(\sqrt{n}))$ and correlations persist over arbitrarily long distances. In contrast to the free boundary case, it has been conjectured that, in the presence of an all-$(+)$ boundary, the mixing time should remain polynomial in n *at all temperatures* [BM02, FH87].

This captures the intuition that the only obstacle to rapid mixing for $\beta > \beta_c$ is the long time required for the dynamics to get through the "bottleneck" between the $(+)$-phase and the $(-)$-phase; the presence of the $(+)$-boundary eliminates the $(-)$-phase and hence the bottleneck. Further support for this intuition comes from the known fact that a certain spatial mixing property holds under the all-$(+)$ boundary condition: specifically, it is known that in this case the correlation between the spins of two sites is exponentially small in the distance between them. Even though this conjecture and the intuition behind it have received a lot of attention in the past decade, obtaining formal arguments to resolve it has proved very elusive. The main obstacle seems to be that the current techniques used for showing polynomial mixing time of the dynamics are all insensitive to the boundary condition, and are thus useful only for scenarios in which the mixing time is polynomial in n *uniformly* in the boundary condition (which is known to be false in the above scenario).

We present what we believe are the first boundary-specific techniques for rapid mixing and prove a strong version of the above conjecture in what is known in statistical physics as the *Bethe approximation*, namely when the lattice $\mathbb{Z}^d$ is replaced by a regular tree. Specifically, we analyze the mixing time of the Glauber dynamics for the Ising model on a tree with $(+)$-boundary condition on its leaves, and show that it remains $O(n \log n)$ at all temperatures. (With a free boundary, the mixing time on a tree is polynomial at all temperatures, but the exponent grows arbitrarily large at low temperatures as $\beta \to \infty$.) Our result continues to hold in the presence of an arbitrary external field, even one that exactly offsets the influence of the large boundary of the tree. This is apparently the first result that quantifies the effect of boundary conditions on the dynamics in an interesting scenario. The proof is based on a new equivalence between appropriate notions of temporal and spatial mixing for systems on trees. The main novelty of this equivalence is that the notions of temporal and spatial mixing it considers are boundary specific. This allows us to use the fact that correlations decay with distance under the $(+)$-boundary (even though they persist under the free boundary) to deduce $O(n \log n)$ mixing time of the Glauber dynamics under the $(+)$-boundary (even though the mixing time is much slower under the free boundary).

We stress that, while the tree is simpler in some respects than $\mathbb{Z}^d$ due to the lack of cycles, in other respects it is more complex: e.g., it exhibits a "double phase transition". Moreover, the Ising model on trees has recently received a lot of attention as the canonical example of a statistical physics model on a "non-amenable" graph (i.e., one whose boundary is of comparable size to its volume) — see, e.g., [KMP01, BRSSZ01, BRZ95, EKPS00, Iof96a,

JS99, Lyo00, ST98].

In addition to the above new equivalence, we also present an improved technique for establishing decay of correlations in the Gibbs distribution for spin systems on trees. Combined with our new implication that deduces rapid mixing from decay of correlations, we get a simple criterion for $O(n \log n)$ mixing time of the Glauber dynamics for systems on trees. We apply this criterion to a number of models other than Ising, including the hard-core model (independent sets), the antiferromagnetic Potts model (colorings), and the (ferromagnetic) Potts model. As a result, we significantly extend the regime of parameters of these models for which the Glauber dynamics is known to mix in $O(n \log n)$ time. This includes extending the regime for which the mixing time is $O(n \log n)$ uniformly in the boundary condition, as well as giving first rapid mixing results for specific boundary conditions.

Our equivalence between temporal and spatial mixing is based on analyzing the *spectral-gap* and *log-Sobolev constant* of the dynamics. These quantities measure the rate of convergence to equilibrium by bounding the rate of decay (in time) of variance and entropy respectively. The main technique we use to analyze the two quantities is decomposing variance and entropy w.r.t. near-product distributions. A similar decomposition approach has been used to prove the (uniform in the boundary condition) equivalence on $\mathbb{Z}^d$ (see, e.g., [Ces01]). We extend this theory of decomposition so that it can be used in boundary-specific scenarios. Since we believe this theory to be central to correspondences between temporal and spatial mixing, we discuss it in a stand-alone appendix where known and new results are collected together, allowing for an accessible study of this theory.

Our new boundary-specific equivalence for trees encourages us to reconsider systems on the integer lattice and look for a similar boundary-specific correspondence. We end this dissertation with an exploration of this direction, where we give a new boundary-specific correspondence for systems on the square integer lattice, based on the decomposition theory mentioned above. However, this correspondence still considers a form of spatial mixing that is stronger than the one that is known to take place in the Ising model at low temperature with a $(+)$-boundary, and thus has no immediate applications. Nevertheless, we hope this new correspondence will motivate further research on this problem.

Organization and bibliographical notes

We give precise definitions of spin systems, Gibbs distributions, Glauber dynamics, and various notions of temporal and spatial mixing in Chapter 2. Our generalizations of the Dobrushin and Dobrushin-Shlosman conditions are given in Chapter 3; the content of this chapter has been submitted for publication as [Wei03]. The combinatorial proofs for the known equivalence between temporal and spatial mixing on the integer lattice are given in Chapter 4; this chapter is based on joint work [DSVW02] with Martin Dyer, Alistair Sinclair and Eric Vigoda. In Chapter 5 we present our new boundary-specific equivalence for systems on trees, as well as its applications; this chapter is based on joint work [MSW03, MSW04] with Fabio Martinelli and Alistair Sinclair. The discussion of boundary-specific mixing on the square integer lattice is given in Chapter 6 and appears only in this dissertation. The theory of decomposing variance and entropy is presented in Appendix A; many of the new proofs in this appendix are taken from [MSW03]. Appendix B contains a proof deferred from Chapter 5 and is also taken from [MSW03].

Chapter 2

Preliminaries

In this chapter we set the basis for the discussion in the rest of this thesis by defining the model of nearest-neighbor spin systems, explaining the concept of a Gibbs measure, and describing the dynamical processes that we analyze. We also present different notions of mixing in space, i.e., types of decay with distance of correlations in the Gibbs distribution, and mixing in time, i.e., types of "fast" convergence to equilibrium of the dynamical process.

2.1 Spin systems

Let $G = (V, E)$ be a countably infinite undirected graph that is locally finite (i.e., of bounded degree). Let $\mathcal{S}$ be a finite alphabet referred to as the *spin space*. A *configuration* is then an element $\sigma \in \Omega := \mathcal{S}^V$, or an assignment of spins to V.

We use the following terminology and notation. Elements of V are called *sites*. Subsets of V are called *regions*, and denoted by upper-case Greek letters. If Ψ is a region, then $\Psi^c := V \setminus \Psi$ and $\partial\Psi := \{x \in \Psi^c \mid \exists y \in \Psi \text{ s.t. } \{x, y\} \in E\}$ is the outer boundary of Ψ. For a configuration σ we write σ_x for the spin at site x under σ, and similarly, σ_Ψ for the configuration on Ψ. When we write "$\sigma = \eta$ on Ψ" we mean that $\sigma_\Psi = \eta_\Psi$. Similarly, "$\sigma = \eta$ off Ψ" means that $\sigma_{\Psi^c} = \eta_{\Psi^c}$. We let $\Omega_\Psi^\eta = \{\sigma \in \Omega \,|\, \sigma = \eta \text{ off } \Psi\}$ denote the set of configurations that agree with the fixed configuration (or "boundary condition") η outside Ψ.

We consider spin systems with nearest neighbor interactions: each edge $\{x, y\} \in E$ is associated with a symmetric pair potential $U_{\{x,y\}} : \mathcal{S} \times \mathcal{S} \to \mathbb{R} \cup \{\infty\}$, and each vertex $x \in V$ is associated with a self potential $U_x : \mathcal{S} \to \mathbb{R} \cup \{\infty\}$. Then, for a finite region Ψ, the

Hamiltonian $H_\Psi : \Omega \to \mathbb{R} \cup \{\infty\}$ is defined as

$$H_\Psi(\sigma) := \sum_{\{x,y\} \in E \,:\, \{x,y\} \cap \Psi \neq \emptyset} U_{\{x,y\}}(\sigma_x, \sigma_y) + \sum_{x \in \Psi} U_x(\sigma_x). \tag{2.1}$$

The value this Hamiltonian assigns can be considered as the contribution to the energy of σ coming from Ψ. Let η specify a boundary condition. The *finite region Gibbs distribution* on Ψ conditioned on η is defined as:

$$\mu_\Psi^\eta(\sigma) := \begin{cases} \frac{1}{Z_\Psi^\eta} \exp(-H_\Psi(\sigma)) & \text{if } \sigma \in \Omega_\Psi^\eta; \\ 0 & \text{otherwise,} \end{cases} \tag{2.2}$$

where Z_Ψ^η is the appropriate normalizing factor. Notice that the distribution on the configurations of Ψ depends only on $\eta_{\partial\Psi}$.

We will sometimes refer to the Gibbs distribution with a "free" boundary condition. By this we mean the distribution resulting from the above definition, except that the first sum in (2.1) is taken only over edges $\{x, y\}$ where both $x \in \Psi$ and $y \in \Psi$, or equivalently, Ψ is disconnected from the rest of the graph. This means that $H_\Psi(\sigma)$ (and therefore the distribution over configurations in Ψ) no longer depends on the configuration on the boundary of Ψ.

We note that for systems with *hard constraints*, i.e., where some potentials may take infinite values (see Examples 2.2 and 2.4 below), it is not necessary that all the finite Gibbs distributions are well-defined. This is the case if for some boundary condition η and some region Ψ, $H_\Psi(\sigma) = \infty$ for all $\sigma \in \Omega_\Psi^\eta$, because then the normalization constant $Z_\Psi^\eta = 0$. However, μ_Ψ^η is guaranteed to be well-defined if η is a *feasible* configuration. A configuration η is said to be feasible if and only if $U_{x,y}(\eta_x, \eta_y)$ and $U_x(\eta_x)$ are finite for every edge $\{x, y\} \in E$ and every site $x \in V$. In this case $Z_\Psi^\eta > 0$ because $H_\Psi(\eta) < \infty$. Notice also that for feasible η, the support of μ_Ψ^η consists only of feasible configurations. In general, only feasible boundary conditions are of interest and it is enough that the Gibbs distributions are well-defined for these boundary conditions. However, some parts of our discussion require that the Gibbs distributions be well-defined for *all* boundary conditions, even non-feasible ones. We therefore say that a spin system is *permissive* if the associated Gibbs distributions are well-defined for all regions Ψ and boundary conditions η, i.e., if for all Ψ and η there exists at least one configuration $\sigma \in \Omega_\Psi^\eta$ for which $H_\Psi(\sigma) < \infty$. Again, notice that only systems with hard constraints may be non-permissive. From here onwards, when a part of the discussion applies only to permissive systems, this is explicitly mentioned.

We now illustrate the above definitions with some concrete examples. The following four spin systems are among the most widely studied in the literature, and will serve as the motivating examples throughout this thesis. In all four, the pair and self potentials defining the system are uniform over edges and vertices respectively.

Example 2.1 Probably the best known spin system is the *Ising* model. In this case, there are two spin values $\mathcal{S} = \{\pm 1\}$. The potentials are $U_{\{x,y\}}(s_1, s_2) = -\beta \cdot s_1 \cdot s_2$ and $U_x(s) = -\beta \cdot h \cdot s$, where $\beta \in \mathbb{R}^+$ is the inverse temperature and $h \in \mathbb{R}$ is the external field. Thus $\Omega = \{\pm 1\}^V$, and the Gibbs distribution μ assigns higher weight to configurations in which many neighboring spins are aligned with one another, as well as to configurations in which many spins agree with the sign of h. These two effects increase with β and with $|h|$ respectively. In particular, at high temperatures (low β) the spins behave almost independently, while at low temperatures (high β) large connected regions of equal spins tend to form.

Example 2.2 Another famous example is the hard-core model (independent sets). This has been used in statistical physics as a model of lattice gases [Geo88], and also in other areas such as the modeling of communication networks [Kel85]. Again there are just two spins $S = \{0, 1\}$, and we refer to a site as *occupied* if it has spin value 1 and *unoccupied* otherwise. A configuration therefore specifies a subset of occupied sites. The potentials are $U_{\{x,y\}}(1,1) = \infty$, $U_{\{x,y\}}(1,0) = U_{\{x,y\}}(0,0) = 1$, and $U_x(s) = -s \cdot \ln \lambda$, where $\lambda \in \mathbb{R}^+$ is the activity parameter. The infinite energy the edge potential assigns to a pair of occupied sites means that there is a hard constraint forbidding two neighboring sites from both being occupied. Thus, in this model, a configuration η is feasible if and only if it specifies an *independent set* in the graph G. Furthermore, the finite Gibbs distributions are over independent sets σ, and $\mu(\sigma) \propto \lambda^{|\sigma|}$, where $|\sigma|$ is the size (i.e., the number of occupied sites) of the independent set σ. In the hard-core model, the activity parameter λ plays a similar role to that of temperature in the Ising model. Specifically, for low values of λ the density of occupied sites is low so the system is not too constrained, and therefore the spins behave almost independently. On the other hand, for large values of λ the high density of occupied sites means that the number of of constraints is significant enough to yield long range order in the system. Finally, note that the hard-core model is permissive for all values of λ, since the configuration in which all the sites in Ψ are unoccupied is assigned positive probability under all boundary conditions.

Example 2.3 A generalization of the Ising model to more than two spin values was introduced by Potts [Pot] and is known as the (ferromagnetic) Potts model. Here $\mathcal{S} = \{1, 2, \ldots, q\}$ and the potentials are $U_{x,y}(s_1, s_2) = -2\beta\delta_{s_1,s_2}$, $U_x(s) = 0$. Thus the spin at each site can take one of q possible values, and the aggregated potential of any configuration depends on the number of adjacent pairs of equal spins. Note that the Ising model is the special case $q = 2$.

Example 2.4 A natural analog to the Ising and Potts models arises by considering *antiferromagnetic* interactions, i.e., interactions in which neighbors with *unequal* spins are favored. This is the so-called antiferromagnetic Potts model. In this model $\mathcal{S} = \{1, 2, \ldots, q\}$, and the potentials are $U_{x,y}(s_1, s_2) = 2\beta\delta_{s_1,s_2}$, $U_x(s) = 0$, where β is the inverse temperature. The most interesting case of this model is when $\beta = \infty$ (i.e., zero temperature), which introduces hard constraints. Thus if we think of the q spin values as colors, a feasible configuration is a *proper coloring* of G, i.e., an assignment of colors to vertices so that no two adjacent vertices receive the same color. The Gibbs distribution is uniform over proper colorings. In this model it is q that provides the parameterization. Notice that the colorings model is permissive if and only if $q > \Delta$, where Δ is the maximum degree of the underlying graph. Finally, we mention that this model has been widely studied not only in statistical physics, but also in computer science because of its connection to graph coloring. See, e.g., [BW02] for an informative account.

2.2 Gibbs measures, uniqueness and mixing in space

It is immediate from the definition that any finite region Gibbs distribution satisfies what are called the "DLR compatibility conditions", namely, for every Ψ, any feasible η and σ that agree off Ψ, and every $\Lambda \subseteq \Psi$,

$$\mu_\Psi^\eta(\,\cdot\,|\sigma_{\Lambda^c}) = \mu_\Lambda^\sigma. \tag{2.3}$$

An immediate consequence is that μ_Ψ^η is *stationary* under μ_Λ. We illustrate what stationarity means with the following two-step process (over configurations on Ψ). In the first step, a configuration σ is chosen according to μ_Ψ^η. In the second step, a configuration is chosen according to μ_Λ^σ. Stationarity means that the resulting distribution of the two-step process is the same as if we only execute the first step, namely, choosing from μ_Ψ^η.

For the discussion below, it is convenient to refer to the collection of all the finite region Gibbs distributions μ_Ψ^η, as Ψ and η vary, as the *specification* μ. Also, when we say "a specification μ", we implicitly refer to an underlying spin system that gives rise to μ. Given a specification μ, it is natural to extend the notion of DLR compatibility to measures on the *infinite* space that are compatible with all the finite distributions.

Definition 2.1 *A probability measure ν over the subset of feasible configurations is called a* Gibbs measure *for the specification μ if, for every finite region Λ and ν-almost every configuration σ,*

$$\nu(\,\cdot\,|\sigma_{\Lambda^c}) = \mu_\Lambda^\sigma.$$

The physical intuition for a Gibbs measure is that it describes a *macroscopic equilibrium*, i.e., all parts of the system are in equilibrium with their boundaries.

It is well known that, for any specification μ derived as above, at least one Gibbs measure always exists. However, several Gibbs measures (or "phases") for a given specification may coexist (see, e.g., [Geo88] or [GHM01] for details and more on Gibbs measures). The question of whether the Gibbs measure is unique or not is central in statistical physics because it corresponds to whether one or more macroscopic equilibria are possible for a given system. The importance of this concept is explained further in Chapter 3, where criteria for uniqueness of the Gibbs measure are given. We note that any Gibbs measure is a convex combination of limits of finite volume Gibbs distributions along appropriate sequences of regions and boundary conditions (see, e.g., [Geo88]). Therefore, the question of whether the Gibbs measure is unique can be translated to that of whether there is asymptotic independence between the configuration on a finite region and a distant boundary configuration. In order to write the above in a formal way, we introduce the following notation. Let ν_1 and ν_2 be two probability measures on Ω, and Λ be a finite region. Then

$$\|\nu_1 - \nu_2\|_\Lambda \;:=\; \max_{A \subseteq \mathcal{S}^\Lambda} |\nu_1(A) - \nu_2(A)|, \tag{2.4}$$

i.e., $\|\nu_1 - \nu_2\|_\Lambda$ is the *total variation distance* between the projections of ν_1 and ν_2 on $\mathcal{S}^\Lambda$. The Gibbs measure for the specification μ is unique if and only if the following condition holds (see, e.g., [Geo88]).

Proposition 2.2 *A specification μ admits a unique Gibbs measure if and only if, for every finite region Λ, there exists an infinite sequence of regions $\Lambda \subset \Psi_1 \subset \Psi_2 \subset \ldots \subset \Psi_\ell \subset \ldots$ that*

covers V (i.e., $\bigcup_{\ell>0} \Psi_\ell = V$), and for any two feasible configurations η and σ,

$$\|\mu^\eta_{\Psi_\ell} - \mu^\sigma_{\Psi_\ell}\|_\Lambda \underset{\ell\to\infty}{\longrightarrow} 0.$$

We observe that uniqueness of the Gibbs measure is a form of what we call in this thesis "mixing in space", i.e., it corresponds to a particular form of asymptotic independence in the equilibrium state between configurations on two distant regions. In the rest of this section, we describe the other spatial mixing notions we consider.

We start with a notion that is very similar to uniqueness of the Gibbs measure, but requires that the decay with distance be exponential. For two regions Λ, Ψ, let $\mathrm{dist}(\Lambda, \Psi)$ stand for the graph distance between the two regions.

Definition 2.3 *We say the specification μ has* weak spatial mixing *if there exist constants C and $\alpha > 0$ such that, for any two regions $\Lambda \subseteq \Psi$, and any pair of boundary configurations η and σ,*

$$\|\mu^\eta_\Psi - \mu^\sigma_\Psi\|_\Lambda \leq C|\Lambda| \exp(-\alpha \cdot \mathrm{dist}(\Lambda, \partial\Psi)).$$

Although the above notion is called "weak", it is easy to see that it implies uniqueness of the Gibbs measure. The reason for the word "weak" is for contrast with the following stronger notion, in which the influence of the boundary condition decays with the distance from the region of disagreement between the two boundary configurations rather than the distance from the boundary as a whole (see Figure 2.1).

Definition 2.4 *We say the specification μ has* strong spatial mixing *if there exist constants C and $\alpha > 0$ such that, for any two regions $\Lambda \subseteq \Psi$, any $\Delta \subseteq \partial\Psi$, and any pair of boundary configurations η and σ that differ only on Δ,*

$$\|\mu^\eta_\Psi - \mu^\sigma_\Psi\|_\Lambda \leq C|\Lambda| \exp(-\alpha \cdot \mathrm{dist}(\Lambda, \Delta)).$$

Remark: In the literature, it is common to define strong spatial mixing as above, but with Δ restricted to be a single site. However, in those references the discussion is limited to the case in which the underlying graph is an integer lattice $\mathbb{Z}^d$, where the single-site definition and ours are equivalent. In order for strong spatial mixing to indeed be stronger than weak spatial mixing on all graphs, our version is necessary. We also mention that the definitions in the literature vary in the sense that different definitions consider different classes of regions Ψ for which the property should hold (for example, on $\mathbb{Z}^d$, Ψ may be restricted to be a regular box). Here we consider the the strongest version by requiring the property to hold in all regions.

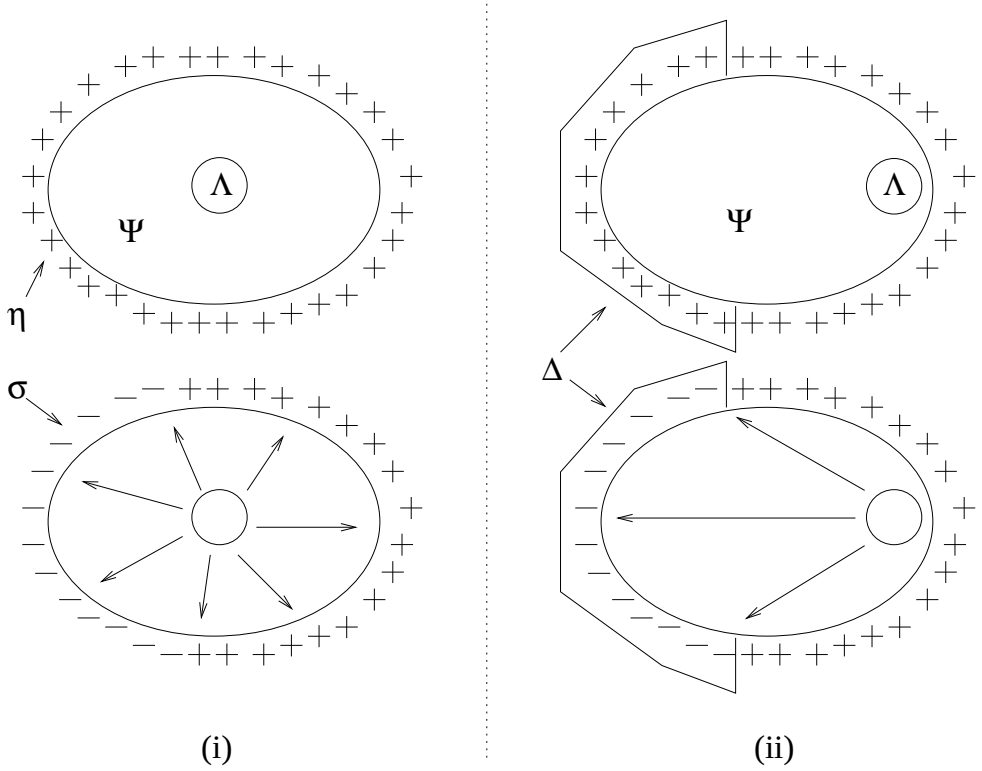

Figure 2.1: Weak (i) vs. strong (ii) spatial mixing. In (i) the total variation distance between the projections on Λ of two Gibbs distributions corresponding to any two boundary conditions η and σ is exponentially small in the graph distance of Λ from $\partial\Psi$. In (ii) it is exponentially small in the distance of Λ from the region of disagreement Δ.

So far, all the spatial mixing notions we have described have compared two different boundary conditions and required that, far inside and away from the boundary, this difference has no effect, i.e., in an appropriate sense, the distribution far from the boundary is roughly the same *uniformly in the boundary condition*. In contrast, the next spatial mixing notion we present considers the Gibbs distribution under a *specific* boundary condition and states that, in this specific distribution, correlations between spins decay with the distance between them. This kind of mixing in space is particularly relevant when there are multiple Gibbs measures, i.e., when different boundary conditions yield significantly different Gibbs distributions. In order to give the precise definition we need to introduce a few more pieces of notation. Let $f : \Omega \to \mathbb{R}$ be a real-valued function on the configuration space. For a boundary condition η and finite region Ψ, we write $\mu^\eta_\Psi(f) = \sum_{\sigma \in \Omega^\eta_\Psi} \mu^\eta_\Psi(\sigma) f(\sigma)$ for the expected value of f w.r.t. μ^η_Ψ. Notice that f is defined on the whole configuration space, and may even depend on the configuration outside Ψ. However, $\mu^\eta_\Psi(f)$ is by definition the expected value of the restriction of f to Ω^η_Ψ, which is consistent with the fact that only configurations in Ω^η_Ψ are in the support of μ^η_Ψ. Continuing with the definitions, let $\mathrm{Var}^\eta_\Psi(f) = \mu^\eta_\Psi(f^2) - \mu^\eta_\Psi(f)^2$ denote the variance of f w.r.t. μ^η_Ψ, and notice that $\mathrm{Var}^\eta_\Psi(f)$ is a measure of the dependence of f on the configuration in Ψ, conditioned on the configuration outside Ψ being η. Similarly, let $\mathrm{Cov}^\eta_\Psi(f,g) = \mu^\eta_\Psi(fg) - \mu^\eta_\Psi(f)\mu^\eta_\Psi(g)$ denote the covariance of f and g w.r.t. μ^η_Ψ. Finally, for a function f that depends only on the configuration on some finite region, let Λ_f denote this region.

Definition 2.5 *Let μ be a specification and η a (boundary) configuration. We say that the Gibbs distribution μ^η_Ψ exhibits an* exponential decay of correlations *if there exists a constant $\alpha > 0$ such that, for any two functions f and g and every Ψ with $\Psi \supseteq \Lambda_f \cup \Lambda_g$,*

$$\mathrm{Cov}^\eta_\Psi(f,g) \leq C_f C_g \exp(-\alpha \cdot \mathrm{dist}(\Lambda_f, \Lambda_g)),$$

where C_f and C_g are constants that depend on f and g respectively.

Remark: In Chapters 5 and 6 we will see that the choice of normalization constants C_f and C_g may play a significant role, i.e., that in some scenarios the above property w.r.t. a certain collection of constants C_f is significantly stronger than the same property w.r.t. larger constants. For now, however, we will be content with this general version of the definition, which is also the common version in the literature.

2.3 The Glauber dynamics and mixing in time

In the previous section we described what an equilibrium state is, and defined various properties which the state may or may not have. In this section, we consider properties of a well known dynamical process that suggests a model for how the system converges to the equilibrium state. This process is called (heat-bath) *Glauber dynamics*, and is a Markov chain Monte Carlo algorithm that is also used for sampling from the Gibbs distribution. A step in this Markov chain is a random update of the spin of a single site conditional on its neighboring spins and, as we will see below, this process indeed converges to the Gibbs distribution.

2.3.1 The Glauber dynamics

In this thesis we consider the Glauber dynamics only on finite regions Ψ, with a fixed (boundary) configuration outside Ψ (although a similar process can also be defined on the infinite graph). The heat-bath Glauber dynamics for a finite region Ψ and boundary condition η is the Markov chain (σ^t) defined as follows. Given the current configuration $\sigma^t \in \Omega_\Psi^\eta$, the transition $\sigma^t \to \sigma^{t+1}$ is defined as:

1. Choose a site x uniformly at random from Ψ.
2. Choose σ^{t+1} from $\mu_{\{x\}}^{\sigma^t}$.

Equivalently, if we let P denote the transition matrix associated with this dynamics (i.e., P is a stochastic matrix whose rows and columns are indexed by configurations in Ω_Ψ^η and such that $P(\sigma_1, \sigma_2)$ is the probability that the chain in σ_1 makes a transition to σ_2) then $P(\sigma_1, \sigma_2) = \frac{1}{|\Psi|} \sum_{x \in \Psi} \mu_{\{x\}}^{\sigma_1}(\sigma_2)$. Notice that unless $\sigma_1 = \sigma_2$, at most one term in the last sum can be positive. This is because $\mu_{\{x\}}^{\sigma_1}(\sigma_2) > 0$ only if $\sigma_1 = \sigma_2$ off x.

We note that for systems with hard constraints there may be issues with the well-definedness of some transitions. If the system is permissive there is no such problem and all transitions are well-defined, even if the current configuration is infeasible. For non-permissive systems, we have to assume that the current configuration is feasible (and in particular, that the fixed boundary condition η is feasible), i.e., the chain is run only on the space of feasible configurations. Notice that in general this is not a problem because only feasible configurations are in the support of μ_Ψ^η.

An easy calculation verifies that the above Markov chain is *reversible* with respect to the Gibbs distribution μ_Ψ^η, i.e., $\mu_\Psi^\eta(\sigma_1)P(\sigma_1,\sigma_2) = \mu_\Psi^\eta(\sigma_2)P(\sigma_2,\sigma_1)$. In particular, μ_Ψ^η is a stationary distribution of the Markov chain. The fact that this is the unique stationary distribution, and that the chain indeed converges to μ_Ψ^η from any initial state, follows from the fact that the chain is aperiodic and connected, (i.e., the graph in which each configuration is a vertex, and $\{\sigma_1,\sigma_2\}$ is an edge if and only if $P(\sigma_1,\sigma_2) > 0$, is connected). To see that the chain is aperiodic, simply observe that $P(\sigma,\sigma) > 0$ (i.e., the chain stays at its current state with positive probability). The state space is guaranteed to be connected if the system has no hard constraints, because in this case $P(\sigma_1,\sigma_2) > 0$ for every pair of configurations σ_1,σ_2 that differ at exactly one site. For systems with hard constraints, we assume that the state space of feasible configurations is connected (for every subset Ψ and boundary condition η), i.e., we limit our discussion of the Glauber dynamics to systems of this kind. Notice that for the hard-core model (Example 2.2), the state space of feasible configurations is connected for all values of λ. For the colorings model (Example 2.4), the state space is connected provided that the number of colors $q \geq \Delta + 2$, where Δ is the maximum degree of a site in G.

Remark: In the literature, the term "Glauber dynamics" usually refers to any Markov chain that makes single-site updates that are reversible with respect to the Gibbs distribution. Our definition is the heat-bath version, where "heat-bath" means that the update at site x is done according the stationary distribution for the spin at x, conditional on the spins of its neighbors. For definiteness, in this thesis "Glauber dynamics" refers to the heat-bath version. Nevertheless, all the results regarding the Glauber dynamics below apply to any version of it.

In some scenarios, it is useful to consider Markov chains that in each step update a finite region (or "block") rather than a single site. The reason for this is that, in some cases, a dynamics that updates larger blocks yields to analysis while the single-site dynamics does not. The analysis of the block dynamics is relevant because it is known that the performance of the single-site dynamics is "similar" to that of the block dynamics, provided that the blocks used in the latter are "not too large". In addition, under the same assumption a step in the chain can be efficiently implemented.

Naturally, in order to define a block dynamics, we have to specify the blocks that may updated. Let $\{\Theta_i\}_{i=1,2,\ldots}$ be a collection of finite regions (blocks) that cover V finitely many times, i.e., each site is included in at least one, but not more than finitely many Θ_i. The heat-bath dynamics for μ_Ψ^η based on the collection of blocks $\{\Theta_i\}_{i=1,2,\ldots}$ is defined as

follows. Let $B(\Psi) := \{i \,|\, \Theta_i \cap \Psi \neq \emptyset\}$ denote the set of indices of blocks that intersect Ψ. The transition $\sigma^t \to \sigma^{t+1}$ is defined as:

1. Choose a block index i uniformly at random from $B(\Psi)$ and let $\Theta = \Theta_i \cap \Psi$.
2. Choose σ^{t+1} from $\mu_\Theta^{\sigma^t}$.

Again, this chain is reversible w.r.t. μ_Ψ^η and converges to it (with the same caveats as before regarding systems with hard constraints). Here, too, one can think of updates other than heat-bath of the chosen block Θ, since all that matters is that the Gibbs distribution μ_Ψ^η be stationary under this update. In Chapter 3 we will consider general types of updates, but elsewhere we limit our attention to the heat-bath version unless otherwise stated. Finally, notice that the Glauber dynamics is the special case in which each block Θ_i is a single site.

2.3.2 Mixing in time

Now that a dynamical process has been defined, and its convergence to the equilibrium distribution established, a natural next step is to quantitatively characterize this approach to equilibrium, i.e., to determine how fast the dynamics converges. From a statistical physics point of view, this question is important for understanding phenomena such as how the system returns to equilibrium after a shock forces it out of it. From an algorithmic point of view, it is important to determine the time it takes for the Markov chain to get close to the Gibbs distribution in order to determine how long the MCMC algorithm should be run in practice.

Different notions of rate of convergence are used and explored in the literature. Probably the most common notion is the *mixing time* $\tau(\epsilon)$, defined as the number of steps required to get within total variation distance ϵ from the stationary distribution, starting from an arbitrary configuration. The formal definition follows. For a region Ψ and boundary condition η, let P be the transition matrix associated with the dynamics for μ_Ψ^η. We write $P^t(\sigma, \cdot)$ for the row indexed by σ in the t-th power of P, and observe that this is the distribution of the dynamics after t steps starting from σ. For two distributions ν_1 and ν_2 whose supports are subsets of Ω_Ψ^η, we write $\|\nu_1 - \nu_2\| = \|\nu_1 - \nu_2\|_\Psi$ for the total variation distance between the two distributions (over the whole of Ψ).

Definition 2.6 *For every $\epsilon > 0$, the mixing time $\tau(\epsilon)$ of the Markov chain P is defined as*

$$\tau(\epsilon) = \inf\left\{t \,:\, \|P^t(\sigma, \cdot) - \mu_\Psi^\eta\| \leq \epsilon \text{ for all } \sigma \text{ in the support of } \mu_\Psi^\eta\right\}.$$

In the literature, the dependency on ϵ is often omitted and the definition only considers a fixed ϵ, e.g., $\epsilon = \frac{1}{2e}$. The reason for this is the following well-known property of the mixing time [Ald83]: for any $\epsilon \leq \frac{1}{e}$, $\tau(\epsilon) \leq \tau(\frac{1}{2e})\lceil \log(\frac{1}{\epsilon}) \rceil$. Nevertheless, we include ϵ in the definition here because, as we will see next, we work with rates of decay in which the dependence on ϵ is sharper than that in the last inequality.

A dual view of the mixing time is the one in which the total variation distance from the stationary distribution is expressed as a function of the number of steps taken (i.e., as a function of time). This view is more convenient for comparing the mixing time with spatial mixing notions, where total variation distance between distributions was expressed as a function of the relevant graph distance (i.e., as a function of space). Before giving the specifics of our temporal mixing notions, we fix two important pieces of notation that are used throughout this thesis. Consider the dynamics based on the collection of blocks $\{\Theta_i\}$ for the region Ψ and some boundary condition η. Let $n = |\Psi|$ denote the volume of the region in which the dynamics takes place, and let $m = |B(\Psi)|$ stand for the number of blocks from which a block to be updated is chosen u.a.r. in every step. (For example, $m = n$ in the case of the single-site Glauber dynamics.)

Definition 2.7 *We say that the dynamics based on the collection of blocks $\{\Theta_i\}$ has* optimal temporal mixing *for a boundary condition η if there exist constants C and $\alpha > 0$ such that, for any region Ψ, the dynamics for μ_Ψ^η has the following property. For any positive integer k and all σ in the support of μ_Ψ^η,*

$$\|P^{km}(\sigma, \cdot) - \mu_\Psi^\eta\| \leq Cn \exp(-\alpha k),$$

where P is the transition matrix associated with the dynamics for μ_Ψ^η.

Remarks:

- Note that optimal temporal mixing here is defined with respect to a specific boundary condition. However, we will sometime discuss this property in a context where it holds *uniformly* in the boundary condition, i.e., for all boundary conditions with uniform constants C and α.
- As in the spatial mixing definitions, other (possibly weaker) versions of the above temporal mixing definition are worth considering, where the property holds for a certain class of regions Ψ rather than for arbitrary regions. We will see an example of this in Chapter 5, where the underlying graph is a regular infinite rooted tree, and where we consider a version of optimal temporal mixing in which the above property holds for all complete subtrees Ψ.

- A simple inversion reveals that optimal temporal mixing is equivalent to a mixing time $\tau(\epsilon) = O(m \log(\frac{n}{\epsilon}))$. Note also that the word *optimal* in the definition should not be taken too literally. Although it is indeed believed that the mixing time of a dynamics that is based on finite size blocks (e.g., the Glauber dynamics) cannot be $o(m \log n)$, we know of no rigorous proof of this conjecture for general spin systems.

In the above temporal mixing notion, when we compare an instance of the chain with the stationary distribution μ_Ψ^η, we measure the total variation distance on the full configuration space, i.e., any subset of $\mathcal{S}^\Psi$ can be the event that distinguishes the two distributions. If optimal temporal mixing holds, then indeed it is guaranteed that after the appropriate number of steps, this total variation distance is small. However, we may also consider the total variation distance between *projections* of the two distributions onto $\mathcal{S}^\Lambda$ for some $\Lambda \subset \Psi$. Obviously, the total variation distance between the projections is at most the total variation distance between the original distributions, but it may be strictly smaller. In particular, it may be the case that better bounds as a function of time can be obtained for the total variation distance between the projections of the distributions on small subsets. This is expressed by the following (stronger) property.

Definition 2.8 *We say that the dynamics based on the collection of blocks $\{\Theta_i\}$ has* optimal projected temporal mixing *for a boundary condition η if there exist constants C and $\alpha > 0$ such that, for any region Ψ, the dynamics for μ_Ψ^η has the following property. For any region $\Lambda \subseteq \Psi$, any positive integer k, and all σ in the support of μ_Ψ^η,*

$$\|P^{km}(\sigma, \cdot) - \mu_\Psi^\eta\|_\Lambda \leq C|\Lambda| \exp(-\alpha k),$$

where P is the transition matrix associated with the dynamics for μ_Ψ^η.

We now pause to discuss the use of the parameters m (the number of blocks) and n (the volume of the region Ψ). First, as already mentioned above, $m = n$ in the case of the Glauber dynamics. In fact, even in the case where the dynamics uses larger blocks, in this thesis the collection of blocks will usually be such that each site is covered by a bounded number of blocks and each block is of bounded size, implying that m is of the same order as n. Thus, if all we were interested in was a bound on the mixing time in terms of the volume n of the region the process is run in, then m could have been replaced by n. However, we retain the parameter m (even in the case of the Glauber dynamics) because we wish to emphasize which parts of our bounds reflect the fact that it takes m steps on average

to update a given block, and which parts are more inherently related to the fact that we sample from a distribution over a region of size n (e.g., because the space of configurations is of size $|\mathcal{S}|^n$). This is especially relevant when considering a continuous time version of the dynamics. Of particular interest is the version where transitions are made m times per unit of time (i.e., each block is updated once per unit of time) on average. This version has an equivalent representation where each block is associated with a mean 1 Poisson clock, and the block is updated whenever its corresponding clock fires (see, e.g., [Mar98]). In fact, when we said that the Glauber dynamics suggests a model for the evolution of the actual physical system towards equilibrium, we were referring to this continuous time version of the dynamics (in which the rate at which a site is updated is independent of the size of the region in which the dynamics takes place). Notice that this version of the Glauber dynamics can be defined not only on finite regions, but on the infinite graph as well. (For precise definitions of the continuous time chain in both the finite and infinite case, see, e.g., [Mar98]). In light of the above discussion, we now illustrate the relevance of the parameter m. For example, "running the dynamics for km steps" should be interpreted as running the dynamics long enough such that each block is updated k times on average; qualitatively (if not precisely), this corresponds to running the continuous time dynamics for k units of time. On the other hand, in "running the dynamics for $km \log n$ steps", the $\log n$ factor should be interpreted as a dependency on n that is inherent to the fact that the dynamics is run on a region of size n, regardless of the rate at which the blocks are updated. In particular, in the continuous time chain described above, this corresponds to running for $k \log n$ units of time.

With the above discussion in mind, it is now easier to see why the notion of optimal projected mixing as in Definition 2.8 is appealing. This is because this notion can be interpreted as a form of bounded mixing time, i.e., one that does not depend on the size n of the region Ψ. To clarify this point, notice that if optimal projected temporal mixing holds then the time needed in order for the projected distribution on a region Λ to be ϵ-close to the stationary distribution is the time such that every block is updated r times on average, where r is a number that depends on ϵ and the size of Λ, but not on the size of the region Ψ in which the dynamics takes place.

We now go on to consider two additional notions of rate of convergence to equilibrium: the *spectral gap* and the *logarithmic Sobolev constant* of the dynamics. These two notions from functional analysis measure the rate of decay with time of *variance* and *en-*

tropy respectively. As we will see below, the mixing time defined above can be bounded in terms of either of the two.

We start with the spectral gap. This is simply the difference between the first and second eigenvalues of the transition matrix P. (Note that the first eigenvalue of this matrix is 1, with μ_Ψ^η as the left eigenvector and the constant vector as the right eigenvector.) Despite this natural definition of the spectral gap, we will work with an equivalent characterization that is more convenient for our purposes. We recall the functional notations $\mu_\Psi^\eta(f)$ and $\mathrm{Var}_\Psi^\eta(f)$, standing for the expected value and variance respectively of the function f w.r.t. μ_Ψ^η. It will be convenient to think of $\mu_\Psi^\eta(f)$ as a function of η, defined by $\mu_\Psi(f)(\eta) = \mu_\Psi^\eta(f)$, the *conditional expectation* of f. In a similar manner, $\mathrm{Var}_\Psi(f)$ is the function representing the conditional variance of f. Let P be the heat-bath Markov chain for the Gibbs distribution μ_Ψ^η based on the collection of blocks $\{\Theta_i\}$. Define the *Dirichlet form* of f w.r.t. P by

$$\tfrac{1}{2}\sum_{\sigma_1,\sigma_2}\mu_\Psi^\eta(\sigma_1)P(\sigma_1,\sigma_2)(f(\sigma_1)-f(\sigma_2))^2 \;=\; \tfrac{1}{m}\sum_{\Theta\in B(\Psi)}\mu_\Psi^\eta(\mathrm{Var}_\Theta(f)) \;\stackrel{\mathrm{def}}{=}\; \tfrac{1}{m}\mathcal{D}_\Psi^\eta(f). \qquad (2.5)$$

(The l.h.s. here is the general definition for any Markov chain; the equality holds when specializing to the case of the heat-bath dynamics, as a simple calculation verifies.) Thus $\mathcal{D}_\Psi^\eta(f)$ is the standard Dirichlet form scaled by a factor of m, and can be thought of as the sum of "local variances" of f in the blocks Θ.

The (scaled) spectral gap compares the sum of local variances $\mathcal{D}_\Psi^\eta(f)$ to the variance of f:

$$c_{\mathrm{gap}} \equiv c_{\mathrm{gap}}(P) \;=\; \inf_f \frac{\mathcal{D}_\Psi^\eta(f)}{\mathrm{Var}_\Psi^\eta(f)} \;=\; \inf_f \frac{\sum_{\Theta\in B(\Psi)}\mu_\Psi^\eta(f)(\mathrm{Var}_\Theta(f))}{\mathrm{Var}_\Psi^\eta(f)}, \qquad (2.6)$$

where the infimum is over non-constant functions f.

We go on to define the log-Sobolev constant. We first extend the functional notation. For a non-negative function $f:\Omega\to\mathbb{R}^+$, let $\mathrm{Ent}_\Psi^\eta(f) = \mu_\Psi^\eta(f\log f) - \mu_\Psi^\eta(f)\log\mu_\Psi^\eta(f)$ denote the entropy of f w.r.t. μ_Ψ^η. Notice that like Var_Ψ^η, $\mathrm{Ent}_\Psi^\eta(f)$ is a (different) measure of the dependency of f on the configuration in Ψ, conditioned on the configuration outside Ψ being η. The (scaled) log-Sobolev constant compares the sum of local variances of $\sqrt{f}$ with the entropy of f:

$$c_{\mathrm{sob}} \equiv c_{\mathrm{sob}}(P) \;=\; \inf_{f\geq 0} \frac{\mathcal{D}_\Psi^\eta(\sqrt{f})}{\mathrm{Ent}_\Psi^\eta(f)}, \qquad (2.7)$$

where the infimum is over non-constant functions f. We note that $c_{\rm sob}$ can be approximately expressed as a ratio of sum of "local entropies" to entropy in the same way that $c_{\rm gap}$ is a ratio of sum of local variances to variance. To see this, we observe that for every region Θ, there exists a constant $\alpha > 0$ (depending only on the potentials of the spin system and the size of Θ) such that $\mathrm{Var}^{\xi}_{\Theta}(\sqrt{f}) \geq \alpha \mathrm{Ent}^{\xi}_{\Theta}(f)$ for every ξ. (It is always the case that $\mathrm{Var}^{\xi}_{\Theta}(\sqrt{f}) \leq \mathrm{Ent}^{\xi}_{\Theta}(f)$.) We therefore conclude that if the blocks Θ are of bounded size then

$$\inf_{f\geq 0} \frac{\sum_{\Theta\in B(\Psi)} \mu^{\eta}_{\Psi}(\mathrm{Ent}_{\Theta}(f))}{\mathrm{Ent}^{\eta}_{\Psi}(f)} \;\geq\; c_{\rm sob} \;\geq\; \alpha \inf_{f\geq 0} \frac{\sum_{\Theta\in B(\Psi)} \mu^{\eta}_{\Psi}(\mathrm{Ent}_{\Theta}(f))}{\mathrm{Ent}^{\eta}_{\Psi}(f)}\,, \tag{2.8}$$

where the constant α described above is w.r.t. the largest (bounded size) block.

We wish to emphasize that the characterization of $c_{\rm gap}$ and $c_{\rm sob}$ given in (2.6) and (2.8) respectively is especially useful for discussing relationships between mixing in time and mixing in space because this characterization expresses both. The temporal interpretation comes from standard bounds on the mixing time in terms of $c_{\rm gap}$ and $c_{\rm sob}$ noted below. On the other hand, the r.h.s. of (2.6) and (2.8) are quantities that depend on the equilibrium state μ^{η}_{Ψ}. Specifically, these quantities measure how well variance (respectively entropy) w.r.t. μ^{η}_{Ψ} can be approximated by the sum of "local" conditional variances (respectively entropies). As we will see in Chapters 5, 6 and in Appendix A, how well the sum of local variances approximates the global variance is intimately related to how close the spins are to being independent, i.e., to what extent the Gibbs distribution mixes in space.

We note that we use a version of the spectral gap and log-Sobolev constant scaled by the number of blocks m because scaled versions are easier to work with when the respective quantities of one dynamics are compared with those of another based on a different set of blocks. Such comparisons are common in our analysis below. In addition, the scaled quantities are in fact the spectral gap and log-Sobolev constant respectively of the continuous time dynamics described before.

To put $c_{\rm gap}$ and $c_{\rm sob}$ in a quantitative context, we note that it is easy to verify that for the Glauber dynamics both are at most 1. This follows by considering the ratios in (2.6) and (2.7) respectively, for a non-constant function f that depends on the spin of a single site. For a dynamics based on a general collection of blocks, both quantities are at most c provided that each site is covered by at most c blocks, as can be seen by considering the same function f as above. However, both $c_{\rm gap}$ and $c_{\rm sob}$ may be much smaller, and in particular, may tend to 0 with the volume n. We say that $c_{\rm gap}$ (respectively $c_{\rm sob}$) for a

boundary condition η is *bounded* if there exists $\alpha > 0$ such that $c_{\mathrm{gap}}(P) \geq \alpha$ (respectively $c_{\mathrm{sob}}(P) \geq \alpha$) for all regions Ψ, where P is the Markov chain for μ_Ψ^η.

As already mentioned above, c_{gap} and c_{sob} give bounds on the rate of decay of variance and entropy respectively, which implies that the mixing time is bounded in terms of these quantities. These classical facts are expressed in the following theorem (see, e.g., [Sal97]).

Theorem 2.9 *Let P be a Markov chain for μ_Ψ^τ, and let $\pi_{\min}$ stand for the minimum non-zero probability of a configuration under μ_Ψ^η. (Notice that $\pi_{\min} \geq \exp(-Cn)$ for some constant C that depends only on the potentials of the spin system and the maximum degree Δ of the graph G.) Then for every $\sigma \in \Omega_\Psi^\eta$,*

(i) $\|P^{km}(\sigma,\cdot) - \mu_\Psi^\eta\| \leq \sqrt{\pi_{\min}^{-1}}\exp(-c_{\mathrm{gap}}k) \leq \exp(Cn - c_{\mathrm{gap}}k)$*, or equivalently, the mixing time* $\tau(\epsilon) \leq c_{\mathrm{gap}}^{-1} \times Cm[n + \log(\frac{1}{\epsilon})]$*;*

(ii) $\|P^{km}(\sigma,\cdot) - \mu_\Psi^\eta\| \leq \log(\pi_{\min}^{-1})\exp(-c_{\mathrm{sob}}k) \leq Cn\exp(-c_{\mathrm{sob}}k)$*, or equivalently,* $\tau(\epsilon) \leq c_{\mathrm{sob}}^{-1} \times m\log(\frac{Cn}{\epsilon})$. □

We conclude this section with a brief comparison of the different temporal mixing notions presented here. An immediate consequence of Theorem 2.9 is that bounded c_{sob} implies optimal temporal mixing. On the other hand, we note that optimal temporal mixing implies that c_{gap} is bounded. The reason for this is that at least for some σ, the variation distance $\|P^{km}(\sigma,\cdot) - \mu_\Psi^\eta\|$ decays (asymptotically) as $\exp(-c_{\mathrm{gap}}k)$. Thus, if optimal temporal mixing holds, then $c_{\mathrm{gap}} \geq \alpha$, where α is the constant in Definition 2.7. We therefore conclude that bounded c_{sob} is stronger then optimal temporal mixing, which is stronger than bounded c_{gap} (though we do not claim the relations to be strict). As for optimal projected temporal mixing, in Chapter 5 we establish examples in which all of the above three notions hold, but optimal projected temporal mixing does not. Thus, optimal projected temporal mixing is strictly stronger than optimal temporal mixing and bounded c_{gap}. We refrain from claiming that optimal projected temporal mixing is strictly stronger than bounded c_{sob} because the two notions may be incomparable, i.e., we are not able to rule out scenarios in which optimal projected temporal mixing holds, but c_{sob} goes to zero with the volume n.

Chapter 3

Criteria for uniqueness of the Gibbs measure

In this chapter we focus on the question of uniqueness of the infinite volume Gibbs measure. As we discussed in Chapter 2, any spin-system admits at least one Gibbs measure; however, a given system may admit multiple Gibbs measures, and one of the central issues in statistical physics is determining whether a spin system admits a unique or multiple Gibbs measures, corresponding to one or more possible equilibrium states. The motivation behind this classification is locating the boundary (in terms of parameters of the system) between the single-phase and multiple-phase regimes. This boundary marks a *phase transition* in the macroscopic behavior of the system, a phenomenon that has additional physical manifestations. For example, the Ising model on the square integer lattice $\mathbb{Z}^2$ with no external field admits a unique Gibbs measure when the temperature is above a known critical value $T_c = 1/\beta_c$, and two distinct Gibbs measures when the temperature is below T_c. One of these Gibbs measures is the limit of μ_Ψ^η as Ψ goes to $\mathbb{Z}^2$, where the boundary configuration η is the all-$(+)$ configuration. The other Gibbs measure is the same limit where η is the all-$(-)$ configuration.

As is apparent from Proposition 2.2, the uniqueness of the Gibbs measure is equivalent to asymptotic independence between the configuration on a finite region and the "boundary" configuration outside a large ball around this region, and thus the phase transition points described above correspond to emergence of long-range correlations (i.e., "order") in the system, or equivalently, to disappearance of the spatial mixing property ex-

pressed in Proposition 2.2. This proposition also explains why discrete mathematicians and probabilists are interested in the subject: the question of uniqueness can be viewed combinatorially as comparing two finite distributions (conditioned on two different boundary configurations), and asking whether or not their difference goes to zero as the boundary ball recedes to infinity.

It is often the case that the Gibbs distributions do not have succinct representations, so that analyzing the asymptotics directly is impossible. Thus, it is important to give *finite* conditions which imply uniqueness of the Gibbs measure. By "finite conditions" we mean conditions that depend only on distributions on regions (blocks) of at most some constant size, and hence can be verified by direct calculation. Dobrushin [Dob70] was the first to give such a condition, which has become widely known as the "Dobrushin Uniqueness Condition." This condition considers only the distributions at single sites. Later, Dobrushin and Shlosman [DS85a] gave a more general condition which may depend on larger blocks (though still of finite size). However, unlike the original Dobrushin condition, their condition is applicable only when the underlying graph of sites is an integer lattice $\mathbb{Z}^d$. Additional versions of the Dobrushin-Shlosman condition were given by others (e.g., Stroock and Zegarlinski [SZ92]), but still only in the context of $\mathbb{Z}^d$. These conditions have turned out to be very useful, since they can be verified by direct calculation for a number of models in appropriate parameter ranges, thus implying uniqueness of the Gibbs measure in a rather straightforward way.

In this chapter, we generalize the above conditions by considering *both* larger blocks *and* any underlying graph. Naturally, all such conditions require that the influence spins at different sites have on each other is "small" in an appropriate sense. However, although they do not mention this explicitly, some of the conditions in the literature require that the total influence *on* a site is small, while others require that the total influence *of* a site is small. We make a clear distinction between these two cases, giving two dual conditions, both of them in the generality described above.

Our proofs are combinatorial in nature and involve a dynamical analysis similar to that carried out for analyzing mixing times of certain dynamics. We make heavy use of couplings, especially the *path coupling* method [BD97]. Our conditions for uniqueness essentially give the stronger property of *weak spatial mixing* (Definition 2.3). In addition, natural extensions of them imply *strong spatial mixing* (Definition 2.4) and *optimal projected temporal mixing* (Definition 2.8) of the corresponding dynamics uniformly in the boundary

configuration. The fact that the question of uniqueness yields to an analysis of a dynamical nature, as well as the fact that the conditions imply mixing in both space and time, is part of the general theme in this thesis of connections between the two types of mixing. (We will present more direct relationships between the two in Chapters 4 and 5.)

We apply our conditions to prove uniqueness of the Gibbs measure (and optimal projected temporal mixing uniformly in the boundary condition) for various models. Although the models we discuss are already known to admit a unique Gibbs measure by other methods, for most of them our results extend the range of parameters for "finite size" conditions of the Dobrushin type can be used to establish uniqueness. In addition, our applications illustrate how our two conditions may be used in different scenarios and clarify the differences between them.

The organization of the rest of the chapter is as follows. In Section 3.1 we give definitions specific to this chapter, background on coupling analysis, and precise statements of our results. Section 3.2 contains the proofs of these theorems. In Section 3.3 we give a few extensions of our results, including their implications for mixing in time. Finally, in Section 3.4 we apply our conditions to various models, thus (re)proving that they admit a unique Gibbs measure.

3.1 Notation and statements of results

In this section we extend some of the definitions given in Chapter 2, introduce notation specific to this chapter, and state our precise results.

3.1.1 Update rules

The conditions we give (and their proofs) are based on notions and tools used in the construction and analysis of local Markov chains similar to those defined in Section 2.3.1. Recall that the definition of a dynamics is based on a collection of blocks $\{\Theta_i\}$, and that the definition in Section 2.3.1 was specific to heat-bath update. Here we extend this definition to consider updates other than heat-bath. However, we still require that the update be "local", i.e., that the result of an update of a block Θ depends only on the configuration on $\Theta \cup \partial\Theta$. (Notice that we allow for dependence on Θ.) Naturally, any local update rule for a specification μ has to be consistent with μ, i.e., the Gibbs distributions have to be sta-

tionary under the update rule. The formal definition of a general local update rule is given below.

Let $\{\Theta_i\}_{i=1,2,\ldots}$ be a collection blocks that cover V finitely many times as in Section 2.3.1. In this chapter, we allow the blocks to have different likelihoods, and thus each Θ_i is assigned a positive weight w_i. (In the corresponding dynamical process, the updated block is chosen at random from some finite subset of the collection, and the probability of choosing Θ_i is proportional to w_i.)

Once a weighted collection of blocks is given, the second ingredient needed in order to complete the specification of an update rule is the collection of distributions that govern the result of an update. Clearly, the distribution over resulting configurations depends on the current configuration. Thus, we need to specify a collection of distributions $\kappa = \{\kappa_i^\tau\}$, indexed by the current configuration τ and the index i of the block to be updated. These distributions have to be "local" and consistent with the Gibbs measure:

Definition 3.1 *We say that κ is a* local update rule *for the specification μ based on the collection of blocks $\{\Theta_i\}$ if $\kappa = \{\kappa_i^\tau\}$ is a collection of probability distributions such that:*

1. *for every configuration τ and every i, κ_i^τ is a probability distribution on $\Omega_{\Theta_i}^\tau$;*
2. *the projections of κ_i^τ and κ_i^σ on $\mathcal{S}^{\Theta_i}$ are the same whenever τ and σ agree on $\Theta_i \cup \partial\Theta_i$, i.e., the distribution κ_i^τ (on the configurations in Θ_i) depends only on $\tau_{\Theta_i \cup \partial\Theta_i}$.*
3. *for every feasible τ and i, $\mu_{\Theta_i}^\tau$ is stationary under κ_i, where the notion of stationarity was explained following equation (2.3);*

Property 1 guarantees that only Θ_i is updated under κ_i^τ while the rest of the configuration remains unchanged. Property 2 expresses the locality requirement, which is natural since the Gibbs distribution on Θ_i is also local, i.e., depends only on $\partial\Theta_i$. Property 3 ensures that the update rule is consistent with the Gibbs distribution[1]. We wish to emphasize the following facts regarding local update rules. First, unlike μ, we always require that κ_i^τ is defined even for infeasible τ. However, the stationarity requirement does not apply to infeasible configurations, and thus, unless τ agrees with some feasible configuration on $\Theta_i \cup \partial\Theta_i$, the specification μ imposes no restriction on the distribution κ_i^τ. Second, the fact that $\mu_{\Theta_i}^\tau$ is stationary under κ_i for every feasible τ implies by (2.3) that, for any $\Psi \supseteq \Theta_i$

[1]Notice that in contrast to the discussion in Section 2.3.1, here we only require that the Gibbs distribution is stationary w.r.t. the update and not that it is the unique stationary distribution.

and every feasible η, μ^{η}_{Ψ} is stationary under κ_i, i.e., any Gibbs distribution in any region that includes Θ_i is unaffected by an update of Θ_i. Third, unlike $\mu^{\tau}_{\Theta_i}$, κ^{τ}_i may depend on the configuration inside Θ_i (as well as the configuration on $\partial\Theta_i$). Nevertheless, a natural choice for κ^{τ}_i is simply $\mu^{\tau}_{\Theta_i}$, which corresponds to the original definition of the heat-bath dynamics in Section 2.3.1. However, other possible and reasonable choices exist. As an example of other possible local update rules, consider a "Metropolis" update where κ^{τ}_i is the distribution resulting from the following process. First, update the configuration in Θ_i by choosing it u.a.r. from $\mathcal{S}^{\Theta_i}$; suppose the resulting configuration is σ. Then, output σ ("accept") with probability $\min\left\{\frac{\exp(-H_{\Theta_i}(\sigma))}{\exp(-H_{\Theta_i}(\tau))}, 1\right\}$ and otherwise output τ ("reject"). There are other examples of more sophisticated update rules which are specific to certain models (e.g., the update rule for proper colorings described in [Vig00], or the one for independent sets [DG00] which we discuss in Section 3.4).

Since (by property 1 of Definition 3.1) the spin of a given site may change only when updating a block that includes this site, we will often need to refer to the subset of such blocks. Recall from Section 2.3.1 that for a region Λ, $B(\Lambda) := \{i \mid \Lambda \cap \Theta_i \neq \emptyset\}$. We write $B(x)$ as shorthand for $B(\{x\})$ and notice that $B(\Lambda) = \bigcup_{x\in\Lambda} B(x)$. Finite subsets of block indices arise throughout our discussion, and for such a subset S we write $w_S := \sum_{i\in S} w_i$ for its aggregated weight. As a final note on blocks, we say that the collection $\{\Theta_i\}$ is of *bounded diameter* if there exists a constant r such that the diameter of any block Θ_i is at most r.

3.1.2 Coupling

A common tool for analyzing Markov chains that use a local update rule is to couple the updates of Θ_i starting from two different configurations. A *coupling* of two distributions ν_1 and ν_2 is any joint distribution whose marginals are ν_1 and ν_2. For any two configurations σ and ξ that differ in exactly one site, let $K_i(\sigma,\xi)$ be a coupling of κ^{σ}_i and κ^{ξ}_i. (These atomic couplings determine a coupling $K(\sigma,\xi)$ for arbitrary pairs of configurations σ,ξ that differ in more than one site, using the path coupling construction explained in Section 3.2.2). If σ and ξ agree on $\Theta_i \cup \partial\Theta_i$, $K_i(\sigma,\xi)$ is always defined as the coupling where the two configurations agree on Θ_i with probability 1. If κ is a local update rule for μ, we call the collection $\{K_i\}$, denoted K, a *coupled update rule* for μ. From here onwards, when we refer to a coupled update rule K, we assume it implicitly specifies the collection of blocks $\{\Theta_i\}$,

their weights $\{w_i\}$ and the local update rule κ according to which K is defined.

Our aim is to give conditions on K that imply uniqueness of the Gibbs measure for the specification μ. Namely, our theorems will be of the form: "If there exists a coupled updated rule K for the specification μ such that K satisfies certain conditions, then there is a unique Gibbs measure that is consistent with μ." The conditions on K will require that under a coupled update, the average "distance" between the two coupled configurations is small. Our notion of distance is specified per site. Let $\rho = \{\rho_x\}_{x \in V}$ be a collection of metrics on the spin space $\mathcal{S}$ (one metric for each site in the graph G). We write $\rho_x(\sigma, \xi)$ for $\rho_x(\sigma_x, \xi_x)$, and abuse this notation when considering a coupling Q by writing $\rho_x(Q)$ for the average distance (w.r.t. the joint distribution Q) between the two coupled configurations. Our notion of distance is extended to regions by summing over single sites, i.e., we let $\rho_\Lambda(\sigma, \xi) := \sum_{x \in \Lambda} \rho_x(\sigma, \xi)$. To illustrate the above notion of distance, we note that in the applications given in Section 3.4 the metrics we use are of the form $\rho_x = u_x \cdot \rho_\delta$, where $u_x \in \mathbb{R}^+$ is a weight associated with the site x and $\rho_\delta(s_1, s_2) = 1$ if $s_1 \neq s_2$ (and naturally, $\rho_\delta(s_1, s_2) = 0$ if $s_1 = s_2$). In this case, $\rho_x(Q)$ is just u_x times the probability that the spins at x differ under the coupling Q, and $\rho_\Lambda(Q)$ is the average weighted *Hamming* distance between the two coupled configurations in Λ.

Our theorems below consider collections of metrics with the following two natural properties. The first property states that the distance at any single site is bounded by a uniform constant: we say that a collection of metrics $\{\rho_x\}$ is *bounded* if

$$\sup_{x \in V} \max_{s_1, s_2 \in \mathcal{S}} \rho_x(s_1, s_2)$$

is finite. The second, stronger property states that the total distance in arbitrarily large regions is bounded by a uniform constant: we say that a collection of metrics $\rho = \{\rho_x\}$ is *summable* if

$$\sum_{x \in V} \max_{s_1, s_2 \in \mathcal{S}} \rho_x(s_1, s_2)$$

is finite.

3.1.3 Results

Once a coupled update rule K and a collection of metrics $\rho = \{\rho_x\}$ are fixed, we are in a position to define the influence of a site y on another site x (w.r.t. K and ρ) in an analogous way to the definition of the "matrix of dependencies" in Dobrushin's condition [Dob70].

Definition 3.2 *For a given coupled update rule K and collection of metrics ρ, define the* influence of site y on site x, *denoted $I_{x \leftarrow y}$, as the smallest constant for which, for all pairs of configurations (σ, ξ) s.t. $\sigma = \xi \operatorname{off} y$,*

$$\sum_{i \in B(x)} w_i \rho_x(K_i(\sigma, \xi)) \;\leq\; \rho_y(\sigma, \xi) I_{x \leftarrow y}.$$

The motivation for the above definition is that $I_{x \leftarrow y} \,/\, w_{B(x)}$ is an upper bound on the average distance between the coupled spins at x (relative to the initial distance between the spins at y) at the end of the following procedure: starting from two configurations that may differ only at y, choose a block $\Theta_i \in B(x)$ with probability $w_i \,/\, w_{B(x)}$ and perform a coupled update of Θ_i. Note that $I_{x \leftarrow y} = 0$ if $y \notin \bigcup_{i \in B(x)} (\Theta_i \cup \partial\Theta_i)$ (i.e., only sites in or adjacent to blocks containing x may have non-zero influence on x). We write $I_{x \leftarrow} := \sum_y I_{x \leftarrow y}$ for the sum of influences of all sites on the site x (and notice by the previous remark that this sum is finite). Our first theorem states that, if the normalized total influence on every site w.r.t. a bounded collection of metrics is less than 1, then the Gibbs measure is unique.

Theorem 3.3 *If a specification μ admits a coupled update rule K together with a bounded collection of metrics ρ for which*

$$\sup_x \left\{ \frac{I_{x \leftarrow}}{w_{B(x)}} \right\} \;<\; 1,$$

then the Gibbs measure for μ is unique; furthermore, if the collection of blocks that K is based on is of bounded-diameter, then μ has weak spatial mixing.

We note that our requirement that the metric collection be bounded is necessary. In our discussion of applications in Section 3.4 we give an example of a specification that admits multiple Gibbs measures but for which there exists a coupled updated rule and an *unbounded* metric collection that satisfy the condition in Theorem 3.3.

Remark: Previously known conditions involving the total influence on a site are the single-site Dobrushin condition [Dob70] and the condition referred to as $DSU(Y)$ by Stroock and Zegarlinski [SZ92]. Both conditions only consider the case in which $\rho_x = \rho_\delta$ for all x, where ρ_δ was defined at the end of Section 3.1.2. In addition, the Dobrushin condition only considers the case in which each Θ_i is a single site. The condition of Stroock and Zegarlinski, while considering blocks of larger size as we do, only considers the special case where the underlying graph G is an integer lattice $\mathbb{Z}^d$. Thus, our Theorem 3.3 is a generalization of both Dobrushin's condition and the Stroock and Zegarlinski one.

In our second theorem we consider a natural *dual* condition, namely, that the total influence *of* every site is small. Following the line established in the previous condition, we write $I_{\leftarrow y}$ for the total influence of site y. Although it might seem natural to define $I_{\leftarrow y}$ as $\sum_x I_{x \leftarrow y}$, the appropriate definition turns out to be a slightly more relaxed one obtained by changing the order of quantification over pairs of configurations:

Definition 3.4 *For a given coupled update rule K and collection of metrics ρ, define the* total influence of site *y, denoted $I_{\leftarrow y}$, as the smallest constant for which, for all pairs of configurations (σ, ξ) s.t. $\sigma = \xi \operatorname{off} y$,*

$$\sum_i w_i \rho_{\Theta_i}(K_i(\sigma, \xi)) \;\leq\; \rho_y(\sigma, \xi) I_{\leftarrow y}.$$

Again, there are only finitely many non-zero terms in the sum since there are only finitely many blocks Θ_i which are affected by y. The relevance of this definition comes from the fact that $I_{\leftarrow y}$ is related to the average total distance resulting from an update of a block randomly chosen from those affected by y, when starting from two configurations that differ only at y. (The exact relationship between $I_{\leftarrow y}$ and this distance is rather involved; the detailed bound is given in Section 3.2). To see the connection to the previous definition of influence, notice that $I_{\leftarrow y} \leq \sum_x I_{x \leftarrow y}$. In fact, the only difference between these two expressions is that in $\sum_x I_{x \leftarrow y}$ the quantification over pairs of configurations is taken separately for each x, while in the definition of $I_{\leftarrow y}$ the quantification is taken once, before summing over x (the summation over x comes from the expansion of ρ_{Θ_i}).

Compared to the condition in Theorem 3.3, our condition for uniqueness based on the influence *of* a site places a stronger restriction on the metric collection we are allowed to use by requiring that it be summable.

Theorem 3.5 *If a specification μ admits a coupled update rule K together with a summable collection of metrics ρ that satisfy $\sup_y w_{B(y)} < \infty$, $\inf_y w_{B(y)} > 0$ and*

$$\sup_y \left\{ \frac{I_{\leftarrow y}}{w_{B(y)}} \right\} \;<\; 1,$$

then the Gibbs measure for μ is unique; furthermore, if the collection of blocks that K is based on is of bounded-diameter, then μ has weak spatial mixing.

Again, the requirement that the metric collection be summable is necessary as is illustrated in Section 3.4, where we also show that the condition $\sup_y w_{B(y)} < \infty$ is necessary. It is

not clear whether the requirement that $\inf_y w_{B(y)} > 0$ is necessary or just an artifact of our proof.

Remarks:

- A previously known condition involving the total influence of a site was given by Dobrushin and Shlosman [DS85a]. However, they only considered the case where the underlying graph G is an integer lattice $\mathbb{Z}^d$ and the collection of blocks $\{\Theta_i\}$ is the set of all translations of some fixed region Θ. In addition, in their condition there is freedom to specify only one metric ρ, so that $\rho_x = \rho$ for all $x \in \mathbb{Z}^d$. Notice that this means, in our language, that the resulting collection of metrics is not summable, which at first sight seems not to fit the framework of Theorem 3.5. However, Theorem 3.5 can still be seen as a generalization of the Dobrushin-Shlosman condition as we now explain. Suppose, as in the Dobrushin-Shlosman setting, that there exists a coupled updated rule K and a single metric ρ for which the condition in Theorem 3.5 holds with $\rho_x = \rho$ for all $x \in \mathbb{Z}^d$, and that the diameter of the blocks Θ_i used by K is bounded by some constant r. We can then construct a slightly modified collection of metrics by letting $\rho'_x = (1+\epsilon)^{-|x|}\rho$, where $|x|$ stands for the distance of the site x from the origin of $\mathbb{Z}^d$ and $\epsilon > 0$ is a small enough constant. Since the volume of a ball around the origin of $\mathbb{Z}^d$ grows subexponentially with the ball's radius, ρ' is clearly summable for any $\epsilon > 0$. On the other hand, it is not too difficult to see that if the condition in Theorem 3.5 holds w.r.t. ρ, and if ϵ is small enough, then the condition also holds with ρ replaced by ρ'. The reason for this is that the influence of a site can increase by a factor of at most $(1+\epsilon)^r$ when replacing ρ by ρ'. In fact, in their proof Dobrushin and Shlosman use a similar construction to the above. Furthermore, the fact that in their condition the metric is the same for all sites restricts their condition to models on $\mathbb{Z}^d$ (or, more precisely, to models on graphs of sub-exponential growth). By allowing different metrics for different sites (but requiring that the collection is summable) we are able to handle arbitrary graphs with no restriction on their geometry.

- At this point it is also worth mentioning that in the literature, the Dobrushin-Shlosman condition is often referred to as a generalization of the single-site Dobrushin condition although in fact the two conditions are dual in nature. The reason for this misconception is that the Dobrushin-Shlosman condition was only stated for translation invariant update rules (for ease of notation), allowing the authors to write it in terms of the total influence *on* a site (or on a block) even though the property they used in the proof is that the total influence *of* a site is small (inequalities 2.24 and 2.26 in the proof of Lemma 2.2 in [DS85a]). To clarify this point further, notice that for specifications on $\mathbb{Z}^d$, when the coupled update rule is translation invariant and the metrics ρ_x are uniform in x, then the matrix of dependencies is translation invariant as well, i.e., $I_{x \leftarrow y}$ depends only on $x - y$ (the difference between the two d-dimensional vectors x and y). Therefore, $\sum_x I_{x \leftarrow y} = \sum_y I_{x \leftarrow y}$ and thus

$\sup_y I_{\leftarrow y} \leq \sup_y \sum_x I_{x \leftarrow y} = \sup_x \sum_y I_{x \leftarrow y} = \sup_x I_{x \leftarrow}$. In other words, in this setting, if the condition involving the total influence on a site holds (Theorem 3.3) then so does the condition involving the total influence of a site (Theorem 3.5).

- In the context of Markov chains, the duality between influence *on* and influence *of* a site was already mentioned in [BD97], where it was referred to as a duality between conditions on the rows and on the columns of the dependency matrix.

- As we will discuss in more detail in section 3.3.2, a strengthening of the conditions in Theorems 3.3 and 3.5 implies that strong spatial mixing holds, and that the dynamics based on the update rule κ has optimal projected temporal mixing uniformly in the boundary condition. This is in fact part of a general relationship between these two notions of mixing that will be discussed in Chapter 4. (We also mention that it is natural to define a continuous time infinite volume dynamics based on the update rule κ, in a similar manner to the infinite volume dynamics mentioned in Section 2.3.2. For such a dynamics boundary conditions do not exist, i.e., there is no problem with sites being close to the boundary, and the conditions given in the theorems here are enough for optimal projected temporal mixing to hold.)

3.2 Proofs

3.2.1 Framework

Our theorems state that, under certain conditions, the Gibbs measure for a given specification μ is unique. Thus, following Proposition 2.2, we will show that if the hypothesis of the theorems is true then, for every finite region Λ, we can find an infinite sequence of finite regions $\{\Lambda_\ell\}_{\ell=0,1,2,\ldots}$ such that $\Lambda = \Lambda_0 \subseteq \Lambda_1 \subseteq \ldots \subseteq \Lambda_\ell \subseteq \ldots$, and for any two (boundary) configurations η and τ, $\|\mu^\eta_{\Lambda_\ell} - \mu^\tau_{\Lambda_\ell}\|_\Lambda \to 0$ as $\ell \to \infty$.

The construction of the sequence $\{\Lambda_\ell\}$ depends on the collection of blocks $\{\Theta_i\}$ used by the coupled update rule given in the hypothesis of the theorems. For a subset of block indices S, let $\Phi(S) := \bigcup_{i \in S}(\Theta_i \cup \partial\Theta_i)$ stand for the region of sites that may influence the result of an update of a block from S. Then the sequence $\{\Lambda_\ell\}$ is defined recursively as $\Lambda_0 = \Lambda$ and $\Lambda_{\ell+1} = \Phi(B(\Lambda_\ell))$ (see Figure 3.1). The important property of this sequence is that, if $x \in \Lambda_\ell$, then all the sites that have non-zero influence on x (via a coupled updated) are included in $\Lambda_{\ell+1}$. Notice also that, since every site is included in at least one block Θ_i, then $\Psi \subseteq \Phi(B(\Psi))$ and therefore $\Lambda_\ell \subseteq \Lambda_{\ell+1}$. It is also easy to see that the sequence $\{\Lambda_\ell\}$ covers V, i.e., that every site $x \in V$ is in some Λ_ℓ.

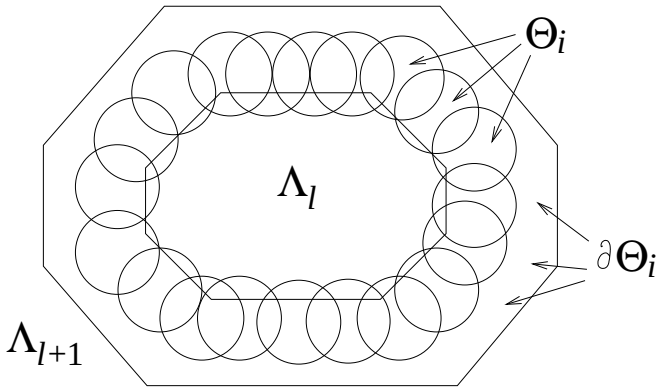

Figure 3.1: Recursively constructing the sequence of regions $\{\Lambda_\ell\}$. The region $\Lambda_{\ell+1}$ contains $\Theta_i \cup \partial\Theta_i$ for each Θ_i that intersects Λ_ℓ.

The proofs of both our theorems will take the following form. For an arbitrary finite region Λ and arbitrary boundary configurations η and τ, using the given coupled update rule we will construct a coupling Q_ℓ of $\mu^\eta_{\Lambda_\ell}$ and $\mu^\tau_{\Lambda_\ell}$ such that $\rho_\Lambda(Q_\ell)$ is exponentially small in ℓ, and in particular, vanishes as ℓ increases. This will conclude the proofs since, if σ and ξ stand for the two coupled configurations under Q_ℓ, then

$$\|\mu^\eta_{\Lambda_\ell} - \mu^\tau_{\Lambda_\ell}\|_\Lambda \;\le\; \Pr{}_{Q_\ell}(\sigma_\Lambda \neq \xi_\Lambda) \;\le\; \frac{\rho_\Lambda(Q_\ell)}{\min_{\sigma_\Lambda \neq \xi_\Lambda} \rho_\Lambda(\sigma_\Lambda, \xi_\Lambda)}\,, \tag{3.1}$$

and $\min_{\sigma_\Lambda \neq \xi_\Lambda} \rho_\Lambda(\sigma_\Lambda, \xi_\Lambda) > 0$ because ρ_Λ is a metric on $\mathcal{S}^\Lambda$. Notice that the weak spatial mixing part of both theorems will follow because if all the blocks in the collection $\{\Theta_i\}$ are of diameter at most r, then Λ_ℓ includes the ball of radius $\ell \cdot r$ around Λ. In particular, for any two regions $\Lambda \subseteq \Psi$, Ψ includes Λ_ℓ for $\ell = \mathrm{dist}(\Lambda, \partial\Psi)/r$. Therefore, from the fact that $\rho_\Lambda(Q_\ell)$ is exponentially small in ℓ and from (3.1), it will follow that $\|\mu^\eta_\Psi - \mu^\tau_\Psi\|_\Lambda$ is exponentially small in $\mathrm{dist}(\Lambda, \partial\Psi)$, as required.

3.2.2 Path coupling

When a coupled update rule K is given then $K_i(\sigma, \xi)$ is specified only for pairs (σ, ξ) that differ in a single site. Based on these atomic couplings, in this subsection we extend this definition to coupled updates for arbitrary pairs of starting configurations. Before doing so, we set notation for an update of a random block. Let S be a finite set of natural numbers indexing blocks. We write $\kappa^\sigma_S := \left(\sum_{i \in S} w_i \kappa^\sigma_i\right) / w_S$ for the distribution resulting from updating a random block from S starting from configuration σ, where the probability of updating Θ_i

for $i \in S$ is proportional to w_i. Similarly, we write $K_S(\sigma,\xi) := \left(\sum_{i\in S} w_i K_i(\sigma,\xi)\right) / w_S$ for a coupled update of a random block from the set S starting from configurations σ and ξ that differ at a single site. Notice that $K_S(\sigma,\xi)$ is indeed a coupling of κ_S^σ and κ_S^ξ.

We now extend the definition of K_S to arbitrary pairs of starting configurations. We first consider pairs (σ,ξ) that agree on $\Phi(S)$ (but may differ in arbitrarily many sites elsewhere). Notice that such pairs induce the same distribution on configurations of $\Phi(S)$ when updating a random block from S, and thus we define $K_S(\sigma,\xi)$ as the trivial coupling where the two resulting configurations agree on $\Phi(S)$ with probability 1. For general σ and ξ, K_S is defined using a *path coupling*. Path couplings (in a more general setting) were introduced in [BD97] where they were used to upper bound the mixing time of certain Markov chains, although similar ideas were already used in the proofs of the uniqueness conditions in [Dob70] and [DS85a].

The idea of a path coupling is to interpolate over differences at single sites, thus reducing the definition of the coupling for general starting pairs (σ,ξ) to those that differ at exactly one site. Although in the literature the interpolation is usually taken only over the sites at which σ and ξ differ, here, in order to ease notation, we interpolate over all sites in $\Phi(S)$. Let $z_1, z_2, \ldots, z_n$ be an enumeration of the sites in $\Phi(S)$, where $n = |\Phi(S)|$. Given σ and ξ, we then construct a sequence of configurations $\sigma^{(0)}, \eta^{(1)}, \ldots, \sigma^{(n)}$ such that $\sigma^{(0)} = \sigma$, and for $1 \le j \le n$, $\sigma_x^{(j)} = \sigma_x^{(j-1)}$ for all $x \neq z_j$ while $\sigma_{z_j}^{(j)} = \xi_{z_j}$. Observe that for every $1 \le j \le n$, $\sigma^{(j)}$ agrees with ξ on $\{z_1,\ldots,z_j\}$ and with σ on $\{z_{j+1},\ldots,z_n\}$. In particular, $\sigma^{(n)}$ agrees with ξ on $\Phi(S)$. Furthermore, $\sigma^{(j-1)}$ and $\sigma^{(j)}$ may only disagree at z_j, and $\sigma^{(j)} = \sigma^{(j-1)}$ if and only if σ and ξ assign the same spin to z_j.

Using the above notation, observe that the couplings $K_S(\sigma^{(j-1)},\sigma^{(j)})$ are already defined for all $1 \le j \le n$, as is the (trivial) coupling $K_S(\sigma^{(n)},\xi)$. We go on to construct the coupling $K_S(\sigma,\xi)$. Recall that $K_S(\sigma,\xi)$ should be a coupling of κ_S^σ and κ_S^ξ, i.e, a coupling of the update of a random block Θ_i, where $i \in S$, starting from σ and ξ respectively. To construct this coupling, first choose a configuration $\tau^{(0)}$ from κ_S^σ. Then, choose a configuration $\tau^{(1)}$ from $\kappa_S^{\sigma^{(1)}}$ according to the coupling $K_S(\sigma,\sigma^{(1)})$ conditioned on $\tau^{(0)}$ being the first configuration in the pair. It is easy to verify that the unconditional distribution of $\tau^{(1)}$ is indeed $\kappa_S^{\sigma^{(1)}}$. Continuing inductively, in step j, choose a configuration $\tau^{(j)}$ from $\kappa_S^{\sigma^{(j)}}$ according to the coupling $K_S(\sigma^{(j-1)},\sigma_j)$ conditioned on $\tau^{(j-1)}$. Finally, choose a configuration $\tau^{(n+1)}$ from κ_S^ξ according to the trivial coupling $K_S(\sigma^{(n)},\xi)$ conditioned on $\tau^{(n)}$. (The last coupling changes the configuration outside $\Phi(S)$ from σ to ξ). Notice that the joint distribution of

$\tau^{(0)}, \tau^{(1)}, \ldots, \tau^{(n+1)}$ is a simultaneous coupling of the distributions $\kappa_S^{\sigma}, \kappa_S^{\sigma^{(1)}}, \ldots, \kappa_S^{\sigma^{(n)}}, \kappa_S^{\xi}$. We define $K_S(\sigma, \xi)$ as the joint distribution of $\tau^{(0)}$ and $\tau^{(n+1)}$, which is indeed a coupling of κ_S^{σ} and κ_S^{ξ}.

The coupling $K_S(\sigma, \xi)$ defined above has the following important property, which can be verified using the triangle inequality for metrics together with the fact that in the above construction the joint distribution of $\tau^{(j-1)}$ and $\tau^{(j)}$ is $K_S(\sigma^{(j-1)}, \sigma^{(j)})$, by definition. For every region $\Delta \subseteq \Phi(S)$,

$$\rho_\Delta(K_S(\sigma, \xi)) \;\leq\; \sum_{j=1}^{n} \rho_\Delta(K_S(\sigma^{(j-1)}, \sigma^{(j)})). \tag{3.2}$$

Now that a coupled update is defined for any two starting configurations, we can define an operator on couplings which, for a given coupling Q, specifies the the result of a coupled update when starting from two configurations chosen from Q.

Definition 3.6 *Let Q be a coupling of two probability distributions ν_1 and ν_2 on Ω. Define*

$$F_S(Q) \;:=\; Q \cdot K_S \;=\; \sum_{\sigma,\xi} Q(\sigma, \xi) K_S(\sigma, \xi),$$

where $Q(\sigma, \xi)$ is the measure of the pair (σ, ξ) under the joint distribution Q. Equivalently, viewing Q as a probability distribution on $\Omega \times \Omega$ and K_S as a Markov kernel on $\Omega \times \Omega$, $F_S(Q)$ stands for the distribution resulting from taking one step in the Markov chain defined by K_S when the starting state is chosen according to Q.

Remark: Even though the space of pairs of configurations is infinite, we used a finite sum notation in Definition 3.6 since in what follows Q will always be a finite distribution, i.e., the support of Q will be a finite subset of pairs of configurations.

Notice that if K is a coupled update rule for μ, and if Q is a coupling of μ_Ψ^η and μ_Ψ^τ for some $\Psi \supseteq \bigcup_{i \in S} \Theta_i$ and any two (boundary) configurations η and τ, then $F_S(Q)$ is a coupling of these two distributions as well. This is because both distributions are stationary under an update of Θ_i for any $i \in S$.

As a final piece of notation, F_S^t stands for t applications of F_S and is the analogue of performing t coupled steps in a Markov chain.

3.2.3 Influence *on* a site

In this subsection we give the proof of Theorem 3.3, namely, that when the influence *on* every site is small, the Gibbs measure is unique. Theorem 3.3 is an immediate consequence

of the following theorem.

Theorem 3.7 *Let μ be a specification, K a coupled updated rule for μ and $\rho = \{\rho_x\}$ a collection of metrics. For any $\delta > 0$, let $\alpha = \delta + \sup_x \left\{ I_{x\leftarrow} \,/\, w_{B(x)} \right\}$, where $I_{x\leftarrow}$ is defined w.r.t. K and ρ. Then, for every finite region Λ, every positive integer ℓ and any two boundary configurations η and τ, there is a coupling Q_ℓ of $\mu^\eta_{\Lambda_\ell}$ and $\mu^\tau_{\Lambda_\ell}$ s.t. $\rho_\Lambda(Q_\ell) \leq c|\Lambda|\alpha^\ell$, where $c = \max_{x\in\Lambda_\ell} \max_{s_1,s_2\in\mathcal{S}} \rho_x(s_1,s_2)$ and the definition of Λ_ℓ is as in Section 3.2.1.*

Notice that if $\sup_x \left\{ I_{x\leftarrow} \,/\, w_{B(x)} \right\} < 1$ as in the hypothesis of Theorem 3.3 then there exists $\delta > 0$ such that $\alpha = \delta + \sup_x I_{x\leftarrow} \,/\, w_{B(x)} < 1$. Furthermore, for a bounded collection of metrics (as in the hypothesis of Theorem 3.3), $c = \max_{x\in\Lambda_\ell} \max_{s_1,s_2\in\mathcal{S}} \rho_x(s_1,s_2)$ is bounded by a constant independent of ℓ. Thus, Theorem 3.3 follows from Theorem 3.7 as explained at the end of Section 3.2.1.

The proof of Theorem 3.7 is based on the following lemma, which for an update of a random block gives an upper bound on the average distance at a site x as a function of the initial distances in the neighborhood of x.

Lemma 3.8 *Fix a coupled updated rule K and a collection of metrics ρ. Let Q be any coupling, x any site and S any finite subset of block indices such that $B(x) \subseteq S$. Then*

$$\rho_x(F_S(Q)) \;\leq\; \left(1 - \frac{w_{B(x)}}{w_S}\right) \rho_x(Q) \;+\; \frac{I_{x\leftarrow}}{w_S} \sup_{y\in\Phi(B(x))} \rho_y(Q). \tag{3.3}$$

Proof: The idea here is that the first term on the r.h.s. of (3.3) represents the contribution to the distance at x when the updated block is not in $B(x)$ (in which case the two spins at x remain unchanged as does the distance at x) while the second term represents the contribution to the distance when the updated block is one from $B(x)$, in which case the distance can be bounded by the total influence on x times the maximum distance of a site that may influence x, as explained below. We proceed with the formal proof. By definition,

$$\rho_x(F_S(Q)) \;=\; \rho_x\left(\sum_{\sigma,\xi} Q(\sigma,\xi)\, K_S(\sigma,\xi)\right) \;=\; \sum_{\sigma,\xi} Q(\sigma,\xi)\, \rho_x(K_S(\sigma,\xi)).$$

We now recall the notation used in the construction of the path coupling in Section 3.2.2, i.e., let $z_1,\ldots,z_n$ enumerate the sites of $\Phi(S)$, where $n = |\Phi(S)|$, and for given σ and ξ let $\sigma = \sigma^{(0)},\ldots,\sigma^{(n)}$ be the corresponding sequence of configurations. Then, using (3.2),

$$\rho_x(F_S(Q)) \;\leq\; \sum_{\eta,\xi} Q(\sigma,\xi) \sum_{j=1}^{n} \rho_x(K_S(\sigma^{(j-1)},\sigma^{(j)})) \;=$$

$$
\begin{aligned}
&= \frac{1}{w_S}\sum_{\sigma,\xi} Q(\sigma,\xi)\sum_{j=1}^{n}\left[\sum_{i\in S\setminus B(x)} w_i\rho_x(K_i(\sigma^{(j-1)},\sigma^{(j)})) + \sum_{i\in B(x)} w_i\rho_x(K_i(\sigma^{(j-1)},\sigma^{(j)}))\right]\\
&= \frac{1}{w_S}\sum_{\sigma,\xi} Q(\sigma,\xi)\left[\rho_x(\sigma,\xi)\sum_{i\in S\setminus B(x)} w_i + \sum_{j=1}^{n}\sum_{i\in B(x)} w_i\rho_x(K_i(\sigma^{(j-1)},\sigma^{(j)}))\right]\\
&= \left(1-\frac{w_{B(x)}}{w_S}\right)\rho_x(Q) + \frac{1}{w_S}\sum_{j=1}^{n}\sum_{\sigma,\xi} Q(\sigma,\xi)\sum_{i\in B(x)} w_i\rho_x(K_i(\sigma^{(j-1)},\sigma^{(j)})),
\end{aligned}
$$

where we made use of the facts that for $i\notin B(x)$, $\rho_x(K_i(\sigma^{(j-1)},\sigma^{(j)}) = \rho_x(\sigma^{(j-1)},\sigma^{(j)})$ and that $\sum_{j=1}^{n}\rho_x(\sigma^{(j-1)},\sigma^{(j)}) = \rho_x(\sigma,\xi)$. What remains to be shown is that

$$
\sum_{j=1}^{n}\sum_{\sigma,\xi} Q(\sigma,\xi)\sum_{i\in B(x)} w_i\rho_x(K_i(\sigma^{(j-1)},\sigma^{(j)})) \;\le\; I_{x\leftarrow}\sup_{y\in\Phi(B(x))}\rho_y(Q). \tag{3.4}
$$

Notice, however, that since $\sigma^{(j-1)}$ and $\sigma^{(j)}$ may differ only at z_j then

$$
\sum_{i\in B(x)} w_i\rho_x(K_i(\sigma^{(j-1)},\sigma^{(j)})) \;\le\; \rho_{z_j}(\sigma^{(j-1)},\sigma^{(j)})I_{x\leftarrow z_j}\,.
$$

Thus, the l.h.s. of (3.4) is bounded by

$$
\begin{aligned}
\sum_{j=1}^{n} I_{x\leftarrow z_j}\sum_{\sigma,\xi} Q(\sigma,\xi)\rho_{z_j}(\sigma^{(j-1)},\sigma^{(j)}) &= \sum_{j=1}^{n} I_{x\leftarrow z_j}\sum_{\sigma,\xi} Q(\sigma,\xi)\rho_{z_j}(\sigma,\xi)\\
&= \sum_{j=1}^{n} I_{x\leftarrow z_j}\rho_{z_j}(Q)\\
&\le \sup_{y\in\Phi(B(x))}\{\rho_y(Q)\}\sum_{y} I_{x\leftarrow y}\\
&= I_{x\leftarrow}\sup_{y\in\Phi(B(x))}\rho_y(Q),
\end{aligned}
$$

where we used the fact that $I_{x\leftarrow y} = 0$ for $y\notin\Phi(B(x))$. □

Lemma 3.8 is useful since it uses only first order information about Q in order to bound $\rho_x(F_S(Q))$, i.e., we only need to know bounds on the average distances at single sites regardless of how these distances depend on each other under Q. In the proof of Theorem 3.7 below, we use Lemma 3.8 iteratively to improve the bounds on single site distances.

Proof of Theorem 3.7: For the δ given in the theorem, let

$$
t_\ell = \left\lceil \frac{w_{B(\Lambda_{\ell-1})}}{\min_{x\in\Lambda_{\ell-1}} w_{B(x)}}\ln\frac{1}{\delta}\right\rceil .
$$

We will show that for any coupling Q, every $0 \leq k \leq \ell$, $x \in \Lambda_{\ell-k}$, and $t \geq kt_\ell$,

$$\rho_x(F^t_{B(\Lambda_{\ell-1})}(Q)) \;\leq\; c\alpha^k, \tag{3.5}$$

where c and α are as defined in Theorem 3.7. The theorem follows from (3.5) as explained next. Take any coupling Q of $\mu^\eta_{\Lambda_\ell}$ and $\mu^\tau_{\Lambda_\ell}$ (for example, the product coupling). Then, for every t, $F^t_{B(\Lambda_{\ell-1})}(Q)$ is also a coupling of $\mu^\eta_{\Lambda_\ell}$ and $\mu^\tau_{\Lambda_\ell}$ because the update rule is consistent with μ and all the blocks that might be updated in the process are included in $\Phi(B(\Lambda_{\ell-1})) = \Lambda_\ell$ by definition. Thus, by setting $t = \ell t_\ell$ we get a coupling Q_ℓ for which $\rho_x(Q_\ell) \leq c\alpha^\ell$ for every $x \in \Lambda_0 = \Lambda$. Hence, $\rho_\Lambda(Q_\ell) \leq c|\Lambda|\alpha^\ell$, as required.

We go on to prove (3.5). Notice that the bound in (3.5) improves as time increases but only when the distance of x from the boundary increases as well, i.e., we only have to consider sites in $\Lambda_{\ell-k}$. The idea of the proof is that once we have established a bound for sites in $\Lambda_{\ell-k}$, we can improve this bound for a site $x \in \Lambda_{\ell-k-1}$ by updating a random block from the ones that cover x, since all the sites that influence x are in $\Lambda_{\ell-k}$. The chosen time parameter ensures that we will indeed update a block from those that cover x with high probability.

The formal proof proceeds by induction on k. The base case ($k = 0$) is clear since $\rho_x(Q) \leq \max_{s_1,s_2\in\mathcal{S}} \rho_x(s_1, s_2) \leq c$ for every $x \in \Lambda_\ell$ by definition of c. We assume (3.5) for k and *show* for $k+1$. Fix an arbitrary $x \in \Lambda_{\ell-k-1}$. We have to show that for every $t \geq (k+1)t_\ell$, $\rho_x(F^t_{B(\Lambda_{\ell-1})}(Q)) \leq c\alpha^{k+1}$. Notice that $y \in \Lambda_{\ell-k}$ for every $y \in \Phi(B(x))$ and hence we can use the induction hypothesis together with Lemma 3.8 to get that, for every $t > kt_\ell$,

$$\rho_x(F^t_{B(\Lambda_{\ell-1})}(Q)) \;\leq\; \left(1 - \frac{w_{B(x)}}{w_{B(\Lambda_{\ell-1})}}\right)\rho_x(F^{t-1}_{\Lambda_{\ell-1}}(Q)) + \frac{I_{x\leftarrow}}{w_{B(\Lambda_{\ell-1})}}c\alpha^k.$$

Therefore,

$$\rho_x(F^t_{B(\Lambda_{\ell-1})}(Q)) - \frac{I_{x\leftarrow}}{w_{B(x)}}c\alpha^k \;\leq\; \left(1 - \frac{w_{B(x)}}{w_{B(\Lambda_{\ell-1})}}\right)\left[\rho_x(F^{t-1}_{B(\Lambda_{\ell-1})}(Q)) - \frac{I_{x\leftarrow}}{w_{B(x)}}c\alpha^k\right]$$

and hence, since by the induction hypothesis $\rho_x(F^{kt_\ell}_{B(\Lambda_{\ell-1})}(Q)) \leq c\alpha^k$, then for all $t \geq kt_\ell$,

$$\rho_x(F^t_{B(\Lambda_{\ell-1})}(Q)) \;\leq\; \frac{I_{x\leftarrow}}{w_{B(x)}}c\alpha^k + \left(1 - \frac{w_{B(x)}}{w_{B(\Lambda_{\ell-1})}}\right)^{t-kt_\ell} c\alpha^k.$$

In particular, for all $t \geq (k+1)t_\ell$,

$$\rho_x(F^t_{B(\Lambda_{\ell-1})}(Q)) \;\leq\; \frac{I_{x\leftarrow}}{w_{B(x)}}c\alpha^k + \delta c\alpha^k \;\leq\; c\alpha^{k+1}.$$

This concludes the proof of (3.5) and thus completes the proof of Theorem 3.7. □

3.2.4 Influence *of* a site

In this section we prove Theorem 3.5, namely, that when the influence *of* every site is small, the Gibbs measure is unique. In contrast to the proof in the previous section, where we used the bound on the influence *on* a site to show that the distance at *every* site decreases exponentially with distance from the boundary, here we will use the bound on the influence *of* a site to show that the *total* distance decreases exponentially with distance. Theorem 3.3 is an immediate consequence of the following theorem.

Theorem 3.9 *Let μ be a specification, K a coupled updated rule for μ and $\rho = \{\rho_x\}$ a collection of metrics. For any $\delta > 0$, let $\alpha = \delta + \sup_y \{I_{\leftarrow y}\} \;/\; \left(\sup_y \{I_{\leftarrow y}\} + \inf_y \left\{w_{B(y)} - I_{\leftarrow y}\right\}\right)$, where $I_{\leftarrow y}$ is defined w.r.t. K and ρ. Then, for every region Λ, any positive integer ℓ and any two boundary configurations η and τ, there is a coupling Q of $\mu^{\eta}_{\Lambda_{\ell+1}}$ and $\mu^{\tau}_{\Lambda_{\ell+1}}$ s.t. $\rho_\Lambda(Q) \le c\alpha^\ell$, where $c = \max_{\sigma,\xi} \rho_{\Lambda_\ell}(\sigma,\xi)$ and Λ_ℓ is as defined in Section 3.2.1.*

Notice that when ρ is summable then the combination of the conditions $\sup_y w_{B(y)} < \infty$, $\inf_y w_{B(y)} > 0$ and $\sup_y \left\{I_{\leftarrow y} \,/\, w_{B(y)}\right\} < 1$ in the hypothesis of Theorem 3.5 is equivalent to the condition $\sup_y \{I_{\leftarrow y}\} \;/\; \left(\sup_y \{I_{\leftarrow y}\} + \inf_y \left\{w_{B(y)} - I_{\leftarrow y}\right\}\right) < 1$. Therefore, for K and ρ as in Theorem 3.5, there exists $\delta > 0$ for which $\alpha < 1$, where α is as defined in Theorem 3.9. Furthermore, the summability of the collection of metrics in Theorem 3.5 implies that $c = \max_{\sigma,\xi} \rho_{\Lambda_\ell}(\sigma,\xi)$ is bounded by a constant independent of ℓ. Thus, Theorem 3.5 follows from Theorem 3.9 as explained at the end of Section 3.2.1.

The proof of Theorem 3.9 is based on the following lemma, which is similar in spirit to Lemma 3.8, but rather than bounding the average distance at a single site, here we bound the average total distance in a region Δ as a function of the initial average total distance in the neighborhood of Δ, when updating of a random block.

Lemma 3.10 *Fix a coupled updated rule K and a collection of metrics ρ. Let Q be any coupling, Δ any region and S any finite subset of block indices such that $B(\Delta) \subseteq S$. Let $\mathrm{MAX} = \max_{y\in\Phi(B(\Delta))} \{I_{\leftarrow y}\}$ and $\mathrm{MIN} = \min_{y\in\Phi(B(\Delta))} \left\{w_{B(y)} - I_{\leftarrow y}\right\}$. Then*

$$\rho_\Delta(F_S(Q)) \;\le\; \left(1 - \frac{\mathrm{MAX} + \mathrm{MIN}}{w_S}\right) \rho_\Delta(Q) \;+\; \frac{\mathrm{MAX}}{w_S} \rho_{\Phi(B(\Delta))}(Q). \tag{3.6}$$

Proof: We start by using the path coupling bound (3.2) to get

$$\rho_\Delta(F_S(Q)) \;\le\; \sum_{\sigma,\xi} Q(\sigma,\xi) \sum_{j=1}^{n} \rho_\Delta(K_S(\sigma^{(j-1)}, \sigma^{(j)})), \tag{3.7}$$

where $n = |\Phi(S)|$ and the sequence of configurations $\sigma^{(j)}$ is as defined in the construction of the path coupling. In turn, we can bound $\rho_\Delta(K_S(\sigma^{(j-1)}, \sigma^{(j)}))$ depending on the location of z_j (the only site at which $\sigma^{(j-1)}$ and $\sigma^{(j)}$ may differ) as follows:

$$\rho_\Delta(K_S(\sigma^{(j-1)}, \sigma^{(j)})) \;\leq\; \rho_{z_j}(\sigma^{(j-1)}, \sigma^{(j)}) \times \begin{cases} I_{\leftarrow z_j} \,/\, w_S + 1 - w_{B(z_j)} \,/\, w_S & \text{if } z_j \in \Delta; \\ I_{\leftarrow z_j} \,/\, w_S & \text{if } z_j \in \Phi(B(\Delta)) \setminus \Delta; \\ 0 & \text{if } z_j \notin \Phi(B(\Delta)). \end{cases} \quad (3.8)$$

Notice that (3.8) follows from the fact that $K_S = (\sum_{i \in S} w_i K_i) \,/\, w_S$, the definition of $I_{\leftarrow z_j}$ and the following four observations. First, $\rho_\Delta(K_i(\sigma^{(j-1)}, \sigma^{(j)})) \leq \rho_{\Theta_i}(K_i(\sigma^{(j-1)}, \sigma^{(j)}))$ if $z_j \in \Theta_i$ because all the sites outside Θ_i remain unchanged and thus the coupled spins of all sites outside Θ_i agree with certainty in the coupling $K_i(\sigma^{(j-1)}, \sigma^{(j)})$. Second, when $z_j \in \Delta \setminus \Theta_i$ then in addition to the distance at Θ_i, there may be positive distance at z_j, which is not accounted for by the distance in Θ_i but which needs to be accounted for as part of the distance in Δ. Thus, in this case, $\rho_\Delta(K_i(\sigma^{(j-1)}, \sigma^{(j)})) \leq \rho_{z_j}(\sigma^{(j-1)}, \sigma^{(j)})) + \rho_{\Theta_i}(K_i(\sigma^{(j-1)}, \sigma^{(j)}))$, where we used the fact that the distance at z_j remains unchanged by the update of Θ_i. Third, when $z_j \notin \Delta$, $\rho_\Delta(K_i(\sigma^{(j-1)}, \sigma^{(j)})) \leq \rho_{\Theta_i}(K_i(\sigma^{(j-1)}, \sigma^{(j)}))$ regardless of whether $z_j \in \Theta_i$ or not because there is no need to count the distance at z_j. Fourth, if $z_j \notin \Phi(B(\Delta))$ then z_j cannot influence the resulting configuration in Δ, i.e., $\rho_\Delta(K_i(\sigma^{(j-1)}, \sigma^{(j)})) = 0$ for all i. This is because the only updates that may incur a non-zero distance at Δ are of blocks for which $\Theta_i \cap \Delta \neq \emptyset$, but then $\sigma^{(j-1)}$ and $\sigma^{(j)}$ agree on $\Theta_i \cup \partial\Theta_i$ since $z_j \notin \Phi(B(\Delta))$ so the distance in Θ_i remains zero.

Now, by plugging the bounds in (3.8) into the r.h.s. of (3.7), and since $\rho_\Delta = \sum_{y \in \Delta} \rho_y$ and $\rho_{z_j}(\sigma^{(j-1)}, \sigma^{(j)})) = \rho_{z_j}(\sigma, \xi)$, we get:

$$\begin{aligned} \rho_\Delta(F_S(Q)) \;&\leq\; \sum_{y \in \Delta} \left(1 - \frac{w_{B(y)}}{w_S} + \frac{I_{\leftarrow y}}{w_S}\right) \rho_y(Q) + \sum_{y \in \Phi(B(\Delta)) \setminus \Delta} \frac{I_{\leftarrow y}}{w_S} \rho_y(Q) \\ &\leq\; \left(1 - \frac{\min_{y \in \Delta} \left\{w_{B(y)} - I_{\leftarrow_y}\right\}}{w_S}\right) \rho_\Delta(Q) + \frac{\max_{y \in \Phi(B(\Delta)) \setminus \Delta} \left\{I_{\leftarrow y}\right\}}{w_S} \rho_{\Phi(B(\Delta)) \setminus \Delta}(Q) \\ &\leq\; \left(1 - \frac{\mathrm{MIN} + \mathrm{MAX}}{w_S} + \frac{\mathrm{MAX}}{w_S}\right) \rho_\Delta(Q) + \frac{\mathrm{MAX}}{w_S} \rho_{\Phi(B(\Delta)) \setminus \Delta}(Q) \\ &=\; \left(1 - \frac{\mathrm{MIN} + \mathrm{MAX}}{w_S}\right) \rho_\Delta(Q) + \frac{\mathrm{MAX}}{w_S} \rho_{\Phi(B(\Delta))}(Q), \end{aligned}$$

where we used the fact that $\rho_\Delta + \rho_{\Phi(B(\Delta)) \setminus \Delta} = \rho_{\Phi(B(\Delta))}$. □

From here onwards the proof of Theorem 3.9 continues in a very similar manner to that of the proof of Theorem 3.7, using Lemma 3.10 iteratively to improve the bounds on average distances in regions.

Proof of Theorem 3.9: For the δ given in the theorem, let

$$t_\ell = \left\lceil \frac{w_{B(\Lambda_\ell)}}{\min_{y\in\Lambda_\ell} w_{B(y)}} \ln \frac{1}{\delta} \right\rceil .$$

We will show that for any coupling Q, every $0 \le k \le \ell$, $x \in \Lambda_{\ell-k}$, and $t \ge kt_\ell$,

$$\rho_{\Lambda_{\ell-k}}(F^t_{B(\Lambda_\ell)}(Q)) \;\le\; c\alpha^k, \tag{3.9}$$

where c and α are as defined in Theorem 3.9. The theorem follows from (3.9) as explained next. Take any coupling Q of $\mu^\eta_{\Lambda_{\ell+1}}$ and $\mu^\tau_{\Lambda_{\ell+1}}$. Then, as we already explained in the proof of Theorem 3.7, for every t, $Q' = F^t_{B(\Lambda_\ell)}(Q)$ is also a coupling of $\mu^\eta_{\Lambda_{\ell+1}}$ and $\mu^\tau_{\Lambda_{\ell+1}}$. Thus, by setting $t = \ell t_\ell$ we get a coupling Q' for which $\rho_{\Lambda_0}(Q') \le c\alpha^\ell$, as required since $\Lambda_0 = \Lambda$.

We go on to prove (3.9). The idea of the proof is that once we have established a bound for the average total distance in $\Lambda_{\ell-k}$, we can improve on this bound for the average total distance in $\Lambda_{\ell-k-1}$ by updating a random block.

The formal proof proceeds by induction on k. The base case ($k = 0$) is clear since $\rho_{\Lambda_\ell}(Q) \le \max_{\sigma,\xi} \rho_{\Lambda_\ell}(\sigma,\xi) \le c$ by definition of c. We assume (3.9) for k and *show* for $k+1$. We have to show that, for every $t \ge (k+1)t_\ell$, we have $\rho_{\Lambda_{\ell-k-1}}(F^t_{B(\Lambda_\ell)}(Q)) \le c\alpha^{k+1}$. Since $\Phi(B(\Lambda_{\ell-k-1})) = \Lambda_{\ell-k}$, we can use the induction hypothesis together with Lemma 3.10 to get that for every $t > kt_\ell$,

$$\rho_{\Lambda_{\ell-k-1}}(F^t_{B(\Lambda_\ell)}(Q)) \;\le\; \left(1 - \frac{\mathrm{MAX}+\mathrm{MIN}}{w_{B(\Lambda_\ell)}}\right) \rho_{\Lambda_{\ell-k-1}}(F^{t-1}_{\Lambda_\ell}(Q)) + \frac{\mathrm{MAX}}{w_{B(\Lambda_\ell)}} c\alpha^k,$$

where $\mathrm{MAX} = \max_{y\in\Lambda_{\ell-k}} \{I_{\leftarrow y}\}$ and $\mathrm{MIN} = \min_{y\in\Lambda_{\ell-k}} \{w_{B(y)} - I_{\leftarrow y}\}$. Therefore,

$$\rho_{\Lambda_{\ell-k-1}}(F^t_{B(\Lambda_\ell)}(Q)) - \frac{\mathrm{MAX}}{\mathrm{MAX}+\mathrm{MIN}} c\alpha^k \;\le$$

$$\left(1 - \frac{\mathrm{MAX}+\mathrm{MIN}}{w_{B(\Lambda_\ell)}}\right) \left[\rho_{\Lambda_{\ell-k-1}}(F^{t-1}_{B(\Lambda_\ell)}(Q)) - \frac{\mathrm{MAX}}{\mathrm{MAX}+\mathrm{MIN}} c\alpha^k\right].$$

Notice that $\min_{y\in\Lambda_{\ell-k}} w_{B(y)} \le \mathrm{MAX}+\mathrm{MIN} \le \max_{y\in\Lambda_{\ell-k}} w_{B(y)}$. In particular, this means that the factor $(1 - \frac{\mathrm{MAX}+\mathrm{MIN}}{w_{B(\Lambda_\ell)}}) \ge 0$. Now, since $\rho_{\Lambda_{\ell-k-1}}(F^{kt_\ell}_{B(\Lambda_\ell)}(Q)) \le \rho_{\Lambda_{\ell-k}}(F^{kt_\ell}_{B(\Lambda_\ell)}(Q)) \le$

$c\alpha^k$ by the induction hypothesis, then for all $t \geq kt_\ell$,

$$\begin{aligned}\rho_{\Lambda_{\ell-k-1}}(F^t_{B(\Lambda_\ell)}(Q)) &\leq \frac{\mathrm{MAX}}{\mathrm{MAX}+\mathrm{MIN}} c\alpha^k + \left(1 - \frac{\mathrm{MAX}+\mathrm{MIN}}{w_{B(\Lambda_\ell)}}\right)^{t-kt_\ell} c\alpha^k \\ &\leq \frac{\mathrm{MAX}}{\mathrm{MAX}+\mathrm{MIN}} c\alpha^k + \left(1 - \frac{\min_{y\in\Lambda_{\ell-k}} w_{B(y)}}{w_{B(\Lambda_\ell)}}\right)^{t-kt_\ell} c\alpha^k.\end{aligned}$$

In particular, for all $t \geq (k+1)t_\ell$,

$$\rho_{\Lambda_{\ell-k-1}}(F^t_{B(\Lambda_\ell)}(Q)) \;\leq\; \frac{\mathrm{MAX}}{\mathrm{MAX}+\mathrm{MIN}} c\alpha^k + \delta c\alpha^k \;\leq\; c\alpha^{k+1}.$$

This concludes the proof of (3.9) and thus completes the proof of Theorem 3.9. □

3.3 Extensions

3.3.1 Extending the model

The conditions in Theorems 3.3 and 3.5 are applicable in more general settings as well. First, the requirement that $\kappa^\tau_{\Theta_i}$ (the result of updating the block Θ_i) depends only on the restriction of τ to $\Theta_i \cup \partial\Theta_i$ can be relaxed to dependency on sites within a bounded radius r from Θ_i. The definition of the sequence $\{\Lambda_\ell\}$ is then adapted to this setting by letting $\Lambda_{\ell+1} = \beta(\Lambda_\ell) \cup \partial_r\beta(\Lambda_\ell)$, where $\partial_r\Lambda$ stands for the set of sites outside Λ that are within distance r from Λ. The rest of the statements and the proofs follow unchanged. Using update rules that depend on sites within distance r is useful when the models have finite range interactions rather than just nearest-neighbor interactions, i.e, potentials are defined for every subset of diameter at most r rather than just single sites and edges, which is the case $r = 1$.

A second observation is that we do not need the spin space $\mathcal{S}$ to be finite, and can instead work with a measurable space $\mathcal{S}$ equipped with a σ-algebra $\mathcal{B}$ of subsets of $\mathcal{S}$. In this case, the metrics ρ_x are required to be measurable functions w.r.t. $\mathcal{B} \times \mathcal{B}$. Up to minor notational and language issues involving infinite spaces, our proofs carry through to this setting except that it may no longer be possible to derive an upper bound on the total variation distance of two distributions when projected onto $\mathcal{S}^\Lambda$ from $\rho_\Lambda(Q)$ as we did in (3.1). However, the rest of our discussion leading to (3.1) is still valid. In particular, under our conditions, for arbitrary τ and σ there exists a coupling of $\mu^\eta_{\Lambda_\ell}$ and $\mu^\tau_{\Lambda_\ell}$ for which $\rho_\Lambda(Q)$ is

exponentially small in ℓ. This means that the Kantorovich-Rubinstein-Ornstein-Vasserstein (KROV) distance (see, e.g., [DS85a] for a definition) between the two distributions w.r.t. ρ_Λ is exponentially small in ℓ and in particular, that the limits of the two sequences of distributions as $\ell \to \infty$ are the same, i.e., the Gibbs measure is unique.

3.3.2 Implications of the conditions for mixing in time

As was shown in [BD97] and [SZ92], conditions similar to the ones in Theorems 3.3 and 3.5 imply optimal temporal mixing of the dynamics based on the relevant update rule. The main difference we have to address in order to apply the conditions to the dynamical setting is that, in the dynamics on a region Ψ, the updated blocks are *intersections* with Ψ of the original blocks Θ_i rather than the Θ_i themselves. This requires us to modify the definition of influence of one site on another such that subsets of Θ_i are considered as well.

To proceed formally, we require here that for every i, the update rule κ is defined for all subsets Θ of Θ_i. (Notice that for heat-bath updates this is automatically given, since the heat-bath rule is defined for every region Θ.) Similarly, the atomic coupling $K_i(\sigma, \xi)$ for σ and ξ that differ at a single site should now be defined for any subset Θ of Θ_i. The influences of and on a site are now defined w.r.t. the region Ψ in which the dynamics takes place, and need only be defined for sites in Ψ. Let $K_i^\Psi(\sigma, \xi)$ denote the coupled update of $\Theta_i \cap \Psi$. Then, for every $x, y \in \Psi$, $I^\Psi_{x \leftarrow y}$ is defined as the smallest constant for which, for all pairs of configurations (σ, ξ) such that $\sigma = \xi \operatorname{off} y$,

$$\sum_{i \in B(x)} w_i \rho_x(K_i^\Psi(\sigma, \xi)) \;\leq\; \rho_y(\sigma, \xi) I^\Psi_{x \leftarrow y},$$

and $I^\Psi_{x \leftarrow} := \sum_{y \in \Psi} I^\Psi_{x \leftarrow y}$. Similarly, $I^\Psi_{\leftarrow y}$ is the smallest constant for which, for all pairs of configurations (σ, ξ) such that $\sigma = \xi \operatorname{off} y$,

$$\sum_i w_i \rho_{(\Theta_i \cap \Psi)}(K_i^\Psi(\sigma, \xi)) \;\leq\; \rho_y(\sigma, \xi) I^\Psi_{\leftarrow y}.$$

Recall the notation m for the number of blocks the dynamics chooses from as discussed in Section 2.3.2. Since in this chapter we allow the blocks to be chosen with different likelihoods, we generalize the notation for the inverse probability of choosing a

block to $m \equiv \lceil \frac{w_{B(\Psi)}}{\min_{i \in B(\Psi)} w_i} \rceil$. Let also

$$\begin{aligned} D_1(\Psi) &= \max_{x \in \Psi, s_1, s_2 \in \mathcal{S}} \rho_x(s_1, s_2); \\ D_2(\Psi) &= \max_{\sigma, \xi} \rho_\Psi(\sigma, \xi); \\ M(\Psi) &= \min_{x \in \Psi, s_1 \neq s_2} \rho_x(s_1, s_2) \;=\; \min_{\sigma_\Psi \neq \xi_\Psi} \rho_\Psi(\sigma_\Psi, \xi_\Psi)\,. \end{aligned}$$

We then have the following two theorems, which are the analogs of Theorems 3.3 and 3.5 respectively.

Theorem 3.11 *Let* $\alpha = \frac{\min_{x \in \Psi} \left\{ w_{B(x)} - I_{x\leftarrow} \right\}}{\min_{i \in B(\Psi)} w_i}$. *If* $\alpha > 0$ *then for every boundary condition* η*, every* $\sigma \in \Omega_\Psi^\eta$*, and any subset* $\Lambda \subseteq \Psi$*,*

$$\| P^{km}(\sigma, \cdot) - \mu_\Psi^\eta \|_\Lambda \;\leq\; \frac{D_1(\Psi)}{M(\Lambda)} |\Lambda| \exp(-\alpha k),$$

where P *is the dynamics for* μ_Ψ^η *based on the update rule* κ*.*

Corollary 3.12 *If the collection of metrics with respect to which* $I_{\leftarrow y}$ *is defined is bounded and* $\sup_\Psi \max_{x \in \Psi} \left\{ I_{x\leftarrow} \,/\, w_{B(x)} \right\} < 1$*, then the dynamics has optimal projected temporal mixing uniformly in the boundary condition.*

Theorem 3.13 *Let* $\alpha = \frac{\min_{y \in \Psi} \left\{ w_{B(y)} - I_{\leftarrow y} \right\}}{\min_{i \in B(\Psi)} w_i}$. *If* $\alpha > 0$ *then for every boundary condition* η*, every* $\sigma \in \Omega_\Psi^\eta$*, and any subset* $\Lambda \subseteq \Psi$*,*

$$\| P^{km}(\sigma, \cdot) - \mu_\Psi^\eta \|_\Lambda \;\leq\; \frac{D_2(\Psi)}{M(\Lambda)} \exp(-\alpha k),$$

where P *is the dynamics for* μ_Ψ^η *based on the update rule* κ*.*

Corollary 3.14 *If the collection of metrics with respect to which* $I_{\leftarrow y}$ *is defined is summable and* $\sup_\Psi \max_{y \in \Psi} \left\{ I_{\leftarrow y} \,/\, w_{B(y)} \right\} < 1$*, then the dynamics has optimal projected temporal mixing uniformly in the boundary condition.*

Remarks:

- Notice that in order to get optimal temporal mixing from Theorems 3.11 and 3.13 respectively, it is enough that $\frac{D_1(\Psi)}{M(\Psi)}$ and $\frac{D_2(\Psi)}{M(\Psi)}$ respectively are polynomial in $n = |\Psi|$. Furthermore, $c_{\text{gap}} \geq \alpha$ regardless of D_1, D_2 and M. (For an explanation of the last fact see the end of Section 2.3.2.)

- Although it is tempting to assert that the influence of y on x through a subset $\Theta \subseteq \Theta_i$ cannot be larger than the influence through Θ_i, this is not the case since the fact that x may be adjacent to the boundary of Θ while being "far" from the boundary of Θ_i may yield a larger influence through Θ (even though y is "far away" in both cases). See [Mar98] for more on this phenomenon of a so-called "boundary phase transition". Nevertheless, it is often the case that bounds given on the influence through Θ_i apply to the influence through any subset Θ of Θ_i. Indeed, in all the applications we give in Section 3.4, we establish not only the conditions for uniqueness given in Theorems 3.3 and 3.5, but in fact the stronger conditions given in Corollaries 3.12 and 3.14.

- Notice that in the infinite volume (continuous time) dynamics that was mentioned in Section 2.3.2 the updates are always of Θ_i in full, and hence it is enough that either of the original conditions in Theorems 3.3 and 3.5 holds in order to get optimal projected temporal mixing of this dynamics.

- In Chapter 4 we show that optimal projected temporal mixing of a dynamics based on bounded-diameter blocks and uniform block weights implies strong spatial mixing. Thus, if either of the conditions in Corollaries 3.12 and 3.14 holds w.r.t. an update rule based on a collection of uniformly weighted bounded-diameter blocks then strong spatial mixing holds. (This can also be proven directly based on arguments similar to those given in the proofs of Theorems 3.3 and 3.5.)

- The fact that stronger conditions such as these indicated here imply optimal temporal mixing and strong spatial mixing was already proved for systems on $\mathbb{Z}^d$ in [DS85b] and [SZ92].

- The converse for systems on $\mathbb{Z}^d$ is also known [DS85b, SZ92]. In particular, if strong spatial mixing holds then the conditions in Corollaries 3.12 and 3.14 hold w.r.t. the heat-bath update of large enough regular boxes. We give another (simple) proof of this fact in Section 4.4.2, where we show that strong spatial mixing implies optimal temporal mixing of the heat-bath dynamics that updates translations of a sufficiently large regular box.

Proof of Theorem 3.11: The proof is based on Lemma 3.8 in a similar way to the proof of Theorem 3.3, i.e., by giving an upper bound on the distance at any site when running a coupled process. For an initial coupling Q, let $F_{B(\Psi)}(Q)$ denote the result of a coupled update of a random block from $B(\Psi)$ as before, but where the update is of $\Theta_i \cap \Psi$ rather than Θ_i. Let Q be any coupling of $P^0(\sigma, \cdot)$ and μ^η_Ψ. (Notice that the first distribution is simply σ, and that the two configurations in the coupling agree outside Ψ with certainty.) Let $\delta_t = \max_x F^t_{B(\Psi)}(Q)$. Then Lemma 3.8 yields $\delta_{t+1} \leq (1 - \frac{\alpha}{m})\delta_t$. Since $\delta_0 \leq D_1(\Psi)$ by definition of D_1, we have $\delta_t \leq D_1 \exp(-\alpha \frac{t}{m})$. Hence, the probability that the two

coupled configurations on Λ differ after km steps is at most $\frac{\delta_{km}}{M(\Lambda)}|\Lambda| \leq \frac{D_1(\Psi)}{M(\Lambda)}|\Lambda|\exp(-\alpha k)$, as required. □

Proof of Theorem 3.13: This theorem is in fact a special case of the general path-coupling argument given in [BD97]. For completeness, we illustrate a proof based on Lemma 3.10 by giving an upper bound on the total distance when running a coupled process. Let Q be any coupling of $P^0(\sigma,\cdot)$ and μ^η_Ψ. Since the two coupled configurations agree with certainty outside Ψ, Lemma 3.10 gives $\rho_\Psi(F_{B(\Psi)}(Q)) \leq (1-\frac{\alpha}{m})\rho_\Psi(Q)$. Since $\rho_\Psi(Q) \leq D_2(\Psi)$ by definition, we get that $\rho_\Lambda(F^t_{B(\Psi)}(Q)) \leq \rho_\Psi(F^t_{B(\Psi)}(Q)) \leq D_2(\Psi)(1-\frac{\alpha}{m})^t$. We conclude that the probability of disagreement in Λ under the coupling after km steps is at most $\frac{\rho_\Lambda(F^{km}_{B(\Psi)}(Q))}{M(\Lambda)} \leq \frac{D_2(\Psi)}{M(\Lambda)}\exp(-\alpha k)$, as required. □

3.4 Applications

In this section we illustrate the use of the conditions given in Theorems 3.3 and 3.5 by carrying out the appropriate calculations for a few specific models in specific ranges of their parameters, thus showing uniqueness of the Gibbs measure for these models in the appropriate ranges. Although we do not extend the previously known range of parameters for which the Gibbs measure is unique, we do extend the range for which finite size conditions of the Dobrushin type hold. In addition, the examples given here shed additional light on our two conditions and the differences between them, and might also serve as guiding examples for readers seeking to establish uniqueness of the Gibbs measure for other models by applying Theorem 3.3 or Theorem 3.5.

The following notation is used in all our examples. Recall that one of the ingredients that needs to be specified in our conditions is a collection of metrics ρ. All the examples we mention in this section use a collection of metrics of the form $\rho_x = u_x\rho_\delta$, where $u_x \in \mathbb{R}^+$ is a weight associated with site x, and ρ_δ is the metric that assigns 1 to any pair of distinct spins and 0 to a pair of identical spins. In particular, for a coupling Q, $\rho_x(Q)$ is exactly u_x times the probability (under Q) that the two coupled spins at x differ. From here onwards a collection of metrics will be specified by determining the set of weights u_x, and implicitly setting $\rho_x = u_x\rho_\delta$.

3.4.1 Colorings of biregular bipartite graphs

We start with an example that emphasizes the differences between the two conditions. These differences are better clarified when the matrix of influences is not symmetric, i.e., the influence of site y on site x is not the same as that of x on y. An example of a model where this symmetry is broken is the model of colorings (as in Example 2.4) on a tree with alternating branching degrees, i.e., the underlying graph is the infinite rooted tree in which vertices at even distance from the root have b_1 children, vertices at odd distance from the root have b_2 children, and $b_1 \neq b_2$. The parameters of the model are thus q (the number of colors) and (b_1, b_2). We apply our conditions to establish that, for $q > b_1 + b_2 + 2$, the Gibbs measure is unique. We give two proofs, one using Theorem 3.3 and the other using Theorem 3.5. We note that uniqueness for colorings on a tree is known to hold for a wider range of parameters. For example, when $b_1 = b_2$ (the tree is regular of degree $b_1 + 1$), the Gibbs measure is unique if and only if $q > b_1 + 1$ [Jon02]. However, the range of parameters for which we show uniqueness here is still larger than that given by the original Dobrushin condition, which is $q > 2(\max\{b_1, b_2\} + 1)$. (The calculation using the original Dobrushin condition was first done in [SS97], and implies uniqueness for $q > 2(b + 1)$ on any underlying graph of maximum degree $b + 1$.

In order to use our theorems we need to specify a coupled update rule for the model. For this example, we use the simple update rule in which each block is a distinct single site, the weights of the blocks are uniform (e.g., all 1) [2], and where updates are done according to the heat-bath rule. We identify a block Θ_i with the site x it consists of and write κ_x^τ in place of κ_i^τ. Notice that for the colorings model, a heat-bath update means that under κ_x^τ, the color at x is chosen uniformly at random from the set of colors not assigned to neighbors of x under τ. Notice also that for $q > \max\{b_1, b_2\} + 1$ (i.e., when the number of colors is larger than the maximum degree of the graph — as is the case for the range of q we consider), κ_x^τ as above is well defined even if τ is infeasible, as required.

In order to complete the specification of our coupled update rule we have to specify how to couple two updates starting from two configurations that disagree at exactly one

[2] When each block is a single site (or more generally, when each site is included in exactly one block), allowing general sets of weights $\{w_i\}$ does not add any generality to our conditions, i.e., for any coupled update rule using a collection of blocks of the above type, the satisfiability of the conditions in Theorems 3.3 and 3.5 is unaffected when changing the set of weights to be uniformly 1. This is because, when each site is covered by exactly one block, the quantity $I_{x\leftarrow} / w_{B(x)}$ in Theorem 3.3 is independent of the choice of weights, and the condition in Theorem 3.5 is not affected if we absorb the weights w_i into the collection of metrics ρ.

site, i.e., we need to specify $K_x(\sigma,\xi)$ for pairs σ and ξ that differ at exactly one site y. We need only specify this coupling when y is a neighbor of x or $y = x$ because otherwise the coupling is required to be the one in which the spins at x agree with probability 1. We set $K_x(\sigma,\xi)$ to be a coupling that minimizes the probability of disagreement between the spins at x. When $y = x$ this simply means the coupling in which the two resulting configurations agree with certainty (since we use a heat-bath update, $\kappa_x^\sigma = \kappa_x^\xi$ when σ and ξ differ only at x). When y is a neighbor of x, and when the number of colors available for the update of the spin at x under σ is the same as under ξ, the optimal coupling is described as follows. Suppose that under both configurations the number of available colors at x is a, and w.l.o.g. that $\sigma_y = 1$ and $\xi_y = 2$. Then, $K_x(\sigma,\xi)$ assigns probability $1/a$ to the pair of configurations in which x is colored 2 and 1 respectively, and for each of the other $a-1$ available colors s, $K_x(\sigma,\xi)$ assigns probability $1/a$ to the pair of configurations in which both spins at x are colored s. Thus, the probability of disagreement is $1/a$. The coupling $K_x(\sigma,\xi)$ takes a similar form when the number of available colors at x under σ is not the same as under ξ (this number may differ by one) so that in either case, the probability of disagreement at x is $1/a$, where a is the number of available colors at x under the configuration for which this number is smaller. In particular, the probability of disagreement is at most $1/(q-\deg(x))$ (with equality for at least one pair σ,ξ). (For details of the calculation when the number of available colors differs, see Section 5.3.4.)

The final ingredient we need to specify is the collection of metrics ρ. Since we use a collection of the form $\rho_x = u_x\rho_\delta$, we only need to specify the weights u_x. Although we use non-uniform weights in order to show uniqueness for the range of parameters mentioned before, it is instructive to first consider the case in which the u_x are uniformly set to 1, which is the setting in the original Dobrushin condition. Under this setting, $I_{x\leftarrow y} = 1/(q-\deg(x))$ since $\rho_x(K_x(\sigma,\xi))$ is simply the probability of disagreement at x under the coupling $K_x(\sigma,\xi)$. Thus, $\sup_x I_{x\leftarrow} = \sup_x \frac{\deg(x)}{q-\deg(x)} = \max\left\{\frac{b_1+1}{q-b_1-1}, \frac{b_2+1}{q-b_2-1}\right\}$. Recall that for the collection of blocks we use, $w_{B(x)} = 1$ for every x since each site is covered by exactly one block whose weight is 1. Thus, using Theorem 3.3 (or equivalently, the original Dobrushin condition) we get that the Gibbs measure is unique in the range of parameters that satisfy $\max\left\{\frac{b_1+1}{q-b_1-1}, \frac{b_2+1}{q-b_2-1}\right\} < 1$, i.e., for $q > 2(\max\{b_1,b_2\}+1)$.

We pause to observe that the colorings model with the above choices of update rule and collection of metrics is a good example of the fact that influence *on* and *of* a site may differ. First, since neighboring sites have different degrees, $I_{x\leftarrow y} \neq I_{y\leftarrow x}$. Furthermore, for a

site y with b_1 children, the total influence of y is $I_{\leftarrow y} = \frac{b_1+1}{q-b_2-1}$. (This is because y has b_1+1 neighbors, and the influence of y on each is $1\,/\,(q-b_2-1)$ because the degree of each neighbor is b_2+1.) Thus, the maximum total influence is $\sup_y I_{\leftarrow y} = \max\left\{\frac{b_1+1}{q-b_2-1}\,,\,\frac{b_2+1}{q-b_1-1}\right\}$, which is less than the maximum total influence on a site when $b_1 \neq b_2$. Notice also that we cannot use the above uniform collection of metrics in order to show uniqueness using the condition based on total influence of a site (Theorem 3.5) because this condition requires that the collection of metrics be bounded, and a uniform collection is clearly not bounded.

We now go on to establish uniqueness of the Gibbs measure for $q > b_1 + b_2 + 2$, improving on the range of parameters for which uniqueness is obtained using the original Dobrushin condition. We give two proofs of this fact, using Theorems 3.3 and 3.5 respectively. For each proof we use a different set of weights u_x. We start with the condition based on the total influence on a site (Theorem 3.3). As is apparent from the analysis of the setting in which u_x is uniform, there is room for improvement since the total influence on sites of the larger degree is larger than on those of the smaller degree. We give the two types of sites different weights in order to balance the total influence they get. Let $u_x = \sqrt{\frac{q-\deg(x)}{\deg(x)}}$. This yields $I_{x\leftarrow y} = \frac{u_x}{u_y(q-\deg(x))} = \sqrt{\frac{\deg(y)}{(q-\deg(y))(q-\deg(x))\deg(x)}}$, and therefore, $I_{x\leftarrow} = \sqrt{\frac{(b_1+1)(b_2+1)}{(q-b_1-1)(q-b_2-1)}}$ for every x. Thus, using Theorem 3.3, the Gibbs measure is unique in the range of parameters for which the last expression is < 1, i.e., for $q > b_1 + b_2 + 2$.

We now give the second proof of uniqueness for the above range of parameters, this time using the condition based on the total influence of a site (Theorem 3.5). In order to use this condition, we have to set the weights u_x so that they yield a summable collection of metrics, i.e., $\sum_x u_x$ has to be finite. In addition, we optimize the weights to minimize the maximum total influence of a site, i.e., we balance the total influence of different sites. In the resulting choice of weights, u_x depends on (and is determined by) the distance of x from the root of the tree. We thus write u_ℓ for the weight of a site at distance ℓ from the root. Set $u_{2\ell} = [(1+\epsilon)b_1b_2]^{-\ell}\sqrt{\frac{q-b_1-1}{b_1+1}}$ and $u_{2\ell+1} = \frac{1}{b_1}[(1+\epsilon)b_1b_2]^{-\ell}\sqrt{\frac{q-b_2-1}{b_2+1}}$, where $\epsilon > 0$ is a small enough constant to be determined later. Clearly, $\sum_x u_x$ is finite because the total weight at level ℓ is proportional to $(1+\epsilon)^{-\lfloor \ell/2 \rfloor}$. We go on to calculate the influence of a site under this choice of weights. Consider a site y at distance 2ℓ from the root. This site influences its parent as well as its b_1 children. The probability of disagreement under the relevant coupling is $1/(q-b_2-1)$ for both the parent and the children, but observe that the

weight of the parent differs from that of the children. Specifically,

$$I_{\leftarrow y} \;=\; \frac{u_{2\ell-1}}{u_{2\ell}(q-b_2-1)} \;+\; b_1\frac{u_{2\ell+1}}{u_{2\ell}(q-b_2-1)}$$

$$=\; \frac{(1+\epsilon)b_2+1}{q-b_2-1}\sqrt{\frac{(b_1+1)(q-b_2-1)}{(q-b_1-1)(b_2+1)}} \;\le\; (1+\epsilon)\sqrt{\frac{(b_1+1)(b_2+1)}{(q-b_1-1)(q-b_2-1)}}\;.$$

For y at distance $2\ell+1$ from the root, a similar calculation gives the slightly better bound $I_{\leftarrow y} \le \sqrt{\frac{(b_1+1)(b_2+1)}{(q-b_1-1)(q-b_2-1)}}$. Applying Theorem 3.5, we conclude that the Gibbs measure is unique in the range of parameters for which $(1+\epsilon)\sqrt{\frac{(b_1+1)(b_2+1)}{(q-b_1-1)(q-b_2-1)}} < 1$ for some $\epsilon > 0$, i.e., for $q > b_1 + b_2 + 2$.

We conclude this subsection by observing that the result obtained here (uniqueness for $q > b_1+b_2+2$) holds for any (b_1+1, b_2+1)-biregular bipartite graph, i.e., for any bipartite graph on vertex set $V_1 \cup V_2$ in which vertices in V_1, V_2 have degrees b_1, b_2 respectively. Indeed, if we examine the first of the two proofs we gave for the tree (the one using total influence *on* a site), we see that the only structure of the graph that we used is that for any site, either its degree is b_1+1 and all its neighbors are of degree b_2+1, or its degree is b_2+1 and all its neighbors are of degree b_1+1. This property holds for any bipartite graph of the above type. We are unaware of any literature discussing colorings of biregular bipartite graphs, and we believe the above bound to be the best known for general graphs of this type (specifically, for those that are not trees). The only previously known bound available for graphs of this type is the one obtained from the original Dobrushin condition (which holds for any graph). As mentioned before, this bound is $q > 2\sup_x \deg(x) = 2(\max\{b_1, b_2\}+1)$, and our bound improves on this for $b_1 \neq b_2$.

3.4.2 Ising model on a regular tree

The next model we discuss is the Ising model (as defined in Example 2.1) on the infinite b-ary tree (in which every vertex except the root has degree $b+1$). Recall that in this model the parameters are the inverse temperature β, the external field h, and the branching degree of the tree b. In this model, the range of parameters for which the Gibbs measure is unique is known exactly [Pre74, Geo88]. Specifically, there exists a critical temperature $\beta_0(b) = \frac{1}{2}\ln(\frac{b+1}{b-1})$ such that for $\beta \le \beta_0$ the Gibbs measure is unique for all external fields h. For $\beta > \beta_0$, there exists a known critical value $h_c(\beta, b) > 0$ such that the Gibbs measure is unique if $|h| > h_c$, and there are multiple Gibbs measures if $|h| \le h_c$. (The phase

diagram of the Ising model on regular trees is discussed in further detail in Section 5.1.2, and illustrated in Figure 5.1.) We will show that both our conditions hold throughout the supercritical regime (i.e., for $\beta > \beta_0$ and arbitrary h, or $\beta \leq \beta_0$ and $|h| > h_c$), an evidence of the tightness of our conditions. We compare this with the range of parameters for which the original Dobrushin condition holds, which is $\beta > \frac{1}{2}\ln(\frac{b+2}{b})$ and arbitrary h. The update rule we use in our proofs is a heat-bath update of a finite subtree. This illustrates another new feature of our conditions, i.e., allowing updates of finite subsets on any graph. As part of our discussion, we also show that the respective restrictions on the collections of metrics in Theorems 3.3 and 3.5 are both necessary. This is done by giving a coupled update rule and collections of metrics that satisfy the conditions in these theorems except for the respective restrictions on the collection of metrics, for some $\beta > \beta_c$ and $h = 0$, i.e., in the regime where there are multiple Gibbs measures.

Single-site updates

We start by considering the heat-bath update rule on single sites and the uniform collection of metrics (where $u_x = 1$ for all x), which is the setting in the original Dobrushin condition. Recall that a heat-bath update simply means that $\kappa_x^\tau = \mu_x^\tau$. For σ and ξ that differ at a single site y, we set $K_x(\sigma, \xi)$ as the optimal coupling of μ_x^σ and μ_x^ξ, i.e., $\rho_x(K_x(\sigma, \xi)) = \|\mu_x^\sigma - \mu_x^\xi\|_x$. Notice that if $y = x$ then $K_x(\sigma, \xi) = 0$ because then μ_x^σ and μ_x^ξ are the same. If y is a neighbor of x, it is well known (e.g., [KMP01]; see also Section 5.3.1) that $\|\mu_x^\sigma - \mu_x^\xi\|_x \leq \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}} \equiv \gamma$ (with equality if the spins of the neighbors of x other than y are divided equally between pluses and minuses). Thus, $I_{x \leftarrow y} \leq \gamma$ (with equality if the degree of x is odd), and the total influence on any site x is $I_{x\leftarrow} \leq (b+1)\gamma$. Using Theorem 3.3 (or, equivalently, the original Dobrushin condition), this immediately establishes uniqueness of the Gibbs measure for β such that $(b+1)\gamma < 1$, i.e., for $\beta < \frac{1}{2}\ln(\frac{b+2}{b})$, and arbitrary h. For the same range of parameters, it is also easy to see that the dual condition in Theorem 3.5 holds for the same coupled update rule, but setting $u_x = [(1+\epsilon)b]^{-|x|}$, where $|x|$ stands for the distance of x from the root of the tree and ϵ is a small enough constant (this is needed in order for the collection of metrics to be summable).

We now use the simple coupled update rule described above in order to show that the restrictions imposed on the collection of metrics used in Theorems 3.3 and 3.5 respectively are necessary. We start with Theorem 3.3. Consider the collection of metrics

resulting from setting $u_x = (\sqrt{b})^{|x|}$. This is clearly not a bounded collection because u_x grows to infinity with the distance of x from the root of the tree. Since now the weight of a site is $\sqrt{b}$ times the weight of its parent, it is easy to see that for every x, $I_{x\leftarrow} \leq [b \cdot \frac{1}{\sqrt{b}} + 1 \cdot \sqrt{b}]\gamma = 2\sqrt{b}\gamma$. Thus, for this choice of weights, $\sup_x I_{x\leftarrow} < 1$ for $\beta < \frac{1}{2}\ln(\frac{2\sqrt{b}+1}{2\sqrt{b}-1})$. However, since $\frac{1}{2}\ln(\frac{2\sqrt{b}+1}{2\sqrt{b}-1}) > \frac{1}{2}\ln(\frac{b+1}{b-1})$ for $b > 4$, for this choice of weights the range of parameters for which the condition holds includes values for which the Gibbs measure is not unique. We thus conclude that the requirement that the collection of metrics be bounded is necessary for Theorem 3.3 to hold.

We go on to consider Theorem 3.5, and show that the requirement that the collection of metrics be summable is necessary for this Theorem to hold. Consider the same coupled update rule as above, but set the weights $u_x = (\sqrt{b})^{-|x|}$. Although this collection is bounded, it is not summable because the total weight of sites at distance ℓ from the root is $(\sqrt{b})^{\ell}$, which goes to infinity with ℓ. A calculation similar to that in the previous paragraph gives that $\sup_y I_{\leftarrow y} < 2\sqrt{b}\gamma$. Hence, as before, for this choice of weights $\sup_y I_{\leftarrow y} < 1$ for some values of β for which the Gibbs measure is not unique. We thus conclude that the requirement that the collection of metrics is summable is necessary in Theorem 3.5.

A slight modification of the last example shows that the requirement that $\sup_y w_{B(y)}$ is finite is also necessary in Theorem 3.5. Recall that so far, we only used collections of blocks in which the weights were uniformly set to 1. In fact, in the rest of the applications in this paper we continue to only use collections of this type, except here, where we wish to demonstrate the necessity of the restriction that $\sup_y w_{B(y)}$ is finite. Thus, consider the coupled update rule from the last two paragraphs, with metric weights $u_x = [(1+\epsilon)b]^{-|x|}$ (so the collection of metrics is summable), except that now the weight of a block $w_x = [(1+\epsilon)\sqrt{b}]^{|x|}$. Since each site y is included in the unique block $\Theta_y = \{y\}$, $B(y)$ consists of only this block and $w_{B(y)} = w_y$. Thus, the above choice of block weights violates the requirement that $\sup_y w_{B(y)}$ is finite because w_y grows to infinity with $|y|$. In addition, it is easy to see that the quantity $\sup_y \left\{ I_{\leftarrow y} \,/\, w_{B(y)} \right\}$ remains exactly the same as in the example of the previous paragraph because the product $w_x u_x$ is unchanged for all x, and since the coupled update rule is the same (up to the change of weights). We thus conclude as in the previous paragraph that the requirement that $\sup_y w_{B(y)}$ is finite is necessary in Theorem 3.5.

Sharp uniqueness bounds using larger blocks

We now go on to show that both our conditions hold throughout the supercritical regime of parameters by considering updates of finite sub-trees.

Theorem 3.15 *In the Ising model on the regular b-ary tree, for the following regimes of parameters there exist a coupled update rule and a collection of metrics that satisfy the condition in Theorem 3.3:*

(i) $\beta < \beta_0$ *and arbitrary* h*;*

(ii) $\beta \geq \beta_0$ *and* $|h| > h_c(\beta)$.

Furthermore, for the same regimes of parameters, there exists a collection of metrics that together with the above coupled update rule satisfies the condition in Theorem 3.5.

Proof: We will only give the proof for regime (i). The proof for regime (ii) goes by a similar but slightly more involved argument, which we sketch at the end. The coupled update rule we use is based on a heat-bath update of a finite size complete subtree. Thus, the collection of blocks is constructed as follows. For every site z, let Θ_z be the complete subtree of height $\ell - 1$ rooted at z, where ℓ is a (large enough) constant to be determined later. Notice that Θ_z consists of ℓ levels (including the level of z itself). The collection of blocks includes Θ_z for every z, plus the $\ell - 1$ blocks which are the complete subtrees of height $0, 1, \ldots, \ell - 2$ respectively, rooted at the root of the original infinite tree (for convenience, we think of these extra blocks as subtrees rooted at imaginary ancestors of the root of the original tree). The addition of the extra blocks guarantees that every site is covered by exactly ℓ blocks. As usual, the weight of every block is set to 1.

As before, we write κ_z^σ for the distribution resulting from an update of Θ_z. Since we use a heat-bath update, $\kappa_z^\sigma = \mu_{\Theta_z}^\sigma$. We need to specify the coupling $K_z(\sigma, \xi)$, where σ and ξ differ at exactly one site $y \in \Theta_z \cup \partial\Theta_z$. If $y \in \Theta_z$ then $K_z(\sigma, \xi)$ is defined as the coupling in which the two configurations agree with certainty. For $y \in \partial\Theta_z$ we use the optimal coupling as constructed in [KMP01], and which is described in detail in Section 5.2.4 (proof of Claim 5.5). In particular, this coupling is constructed recursively along paths of the tree, such that for every $x \in \Theta_z$, the probability of disagreement at x under $K_z(\sigma, \xi)$ is $\leq \gamma^{|x-y|}$, where $|x - y|$ stands for the graph distance between x and y.

Consider now the uniform collection of metrics where $u_x = 1$ for every x. We note that for this choice of weights, if $x \in \Theta_z$ then $\rho_x(K_z(\sigma,\xi)) \leq \gamma^{|x-y|}$, for σ and ξ that disagree only at $y \in \partial\Theta_z$. We go on to calculate the influence a site y has on site x. First, observe that if $|y-x| > \ell$ then $I_{x\leftarrow y} = 0$. If $|y-x| \leq \ell$ (and $y \neq x$) then there is exactly one block Θ_z through which y influences x. This is because y is on the boundary of exactly $b+1$ blocks, namely the b blocks rooted at each of the children of y, and the block rooted at the ancestor ℓ levels above y. Thus, if x is an ancestor of y, then y influences x through Θ_z if and only if z is the ancestor ℓ levels above y. If x is a descendent of y, then y influences x through Θ_z if and only if z is the child of y that is the ancestor of x (and $z = x$ if x is an immediate child of y). We conclude that $I_{x\leftarrow y} \leq \gamma^{|x-y|}$. Thus, for every x,

$$I_{x\leftarrow} \;=\; \sum_y I_{x\leftarrow y} \;\leq\; \frac{b+1}{b}\sum_{j=1}^{\ell}(b\gamma)^j \;.$$

Now, since each site is included in exactly ℓ blocks then $w_{B(x)} = \ell$ for every x. Thus, the above coupled update rule satisfies the condition in Theorem 3.3 if $I_{x\leftarrow} < \ell$ for every x. However, for $\beta < \frac{1}{2}\ln(\frac{b+1}{b-1})$, $\gamma \equiv \frac{e^{\beta}-e^{-\beta}}{e^{\beta}+e^{-\beta}} < \frac{1}{b}$. Thus, for this range of β, $I_{x\leftarrow}$ is bounded by a constant independent of ℓ, and hence $I_{x\leftarrow} < \ell$ for a large enough ℓ (depending on β), as required.

We go on to show that for the same range of parameters of the Ising model, there exists a collection of metrics which together with the above coupled update rule satisfies the condition in Theorem 3.5. Here we need to have a summable collection of metrics, and for this purpose we set $u_x = [(1+\epsilon)b]^{-|x|}$, where $\epsilon > 0$ is a small enough constant (which may depend on (β, b) but not on ℓ) to be set later. Let us calculate $I_{\leftarrow y}$ for this collection of metrics. As before, it is enough to show that $I_{\leftarrow y}$ is bounded by a constant independent of ℓ. First, notice that for the above choice of weights, the total weight of the sites below y is at most a constant times u_y, so the contribution to $I_{\leftarrow y}$ of the blocks immediately below y is bounded by a constant even if the spins of all sites included in these blocks disagree with certainty. We still need to consider the block above y. Let z be the ancestor ℓ levels above y. We need to show that $\rho_{\Theta_z}(K_z(\sigma,\xi))$ is bounded by u_y times a constant independent of ℓ, for every σ and ξ that differ only at y. Since the coupling we use was constructed recursively (see Section 5.2.4), a disagreement at a site $x \in \Theta_z$ can occur only if all the sites on the path from y to x have disagreeing spins. Combining this with the fact that for every site x, the total weight of sites below x is at most a constant times u_x,

we see that $\rho_{\Theta_z}(K_z(\sigma,\xi))$ is bounded by a constant times the average total distance along the path from y to z. It is therefore enough to show that the average total distance along this path is at most a constant (independent of ℓ) times u_y. Notice that the weight of an ancestor j levels above y is $u_y[(1+\epsilon)b]^j$, but that the probability of disagreement at that site under $K_z(\sigma,\xi)$ is at most γ^j. Thus, the above distance along the path is at most $u_y \sum_{j=1}^{\ell}[(1+\epsilon)b\gamma]^j$, which is bounded by a constant times u_y if $\beta < \frac{1}{2}\ln(\frac{b+1}{b-1})$ and ϵ is small enough (such that $(1+\epsilon)\gamma = (1+\epsilon)\frac{e^{\beta}-e^{-\beta}}{e^{\beta}+e^{-\beta}} < \frac{1}{b}$), as required.

We conclude with a few comments about the proof for regime (ii). Notice that in the proof for regime (i) (for both conditions) the crucial properties we used were that for σ and ξ that differ only at y, the probability of disagreement at $x \in \Theta_z$ under $K_z(\sigma,\xi))$ is at most $\gamma^{|x-y|}$, and that $\gamma < \frac{1}{b}$ for β as in regime (i). In regime (ii), the latter bound no longer holds. However, as we show in Section 5.3.1, for the supercritical values of the parameters as in regime (ii), there exist constants c and $\widehat{\gamma} < \frac{1}{b}$ such that the above probability of disagreement is at most $c\widehat{\gamma}^{|x-y|}$. □

Remark: The fact that our conditions hold throughout the uniqueness regime (except at the critical point) is not specific to the Ising model and holds for a number of other models on the regular b-ary tree. In particular, the heat-bath update of complete subtrees of height $\ell - 1$ for large enough ℓ satisfies our conditions throughout the super-critical regime of a number of other models on a regular tree. These include the hard-core model, the colorings model and the ferromagnetic Potts model. This follows from the fact that the quantity $\widehat{\gamma} < \frac{1}{b}$ in these settings. See Section 5.3 for an exact definition of this quantity and the calculation of it in the above settings.

3.4.3 Independent sets of graphs of subexponential growth

In this subsection we discuss an update rule for the hard-core (independent sets) model (defined in Example 2.2) that was presented and analyzed in [DG00] in the dynamical context, where it was shown to have optimal temporal mixing uniformly in the boundary condition $\lambda < \frac{2}{b-1}$, where $b+1$ is the maximum degree of the underlying graph. We put this analysis in the context of our conditions, showing that they are satisfied by the above update rule for the same range of parameters if the underlying graph is of *subexponential growth*. The fact that the Gibbs measure is unique for this range of parameters on graphs of subexponential growth is not new because there is an independent argument (which we discuss in Chapter 4) that states that if a model on a graph of subexponential growth admits

a dynamics that uses bounded diameter blocks (as does the one in [DG00]) and has optimal temporal mixing uniformly in the boundary condition, then strong spatial mixing holds and the Gibbs measure is unique. Our motivation for discussing this update rule is twofold. First, it is an example of an interesting update rule that is more sophisticated than heat-bath. Second, it illustrates how an analysis that was carried out in the dynamical context in order to establish optimal temporal mixing can also be used in order to show that our conditions hold w.r.t. the same update rule and choice of parameters (but only for graphs of subexponential growth).

Before going on to our analysis, we mention some other known bounds for the hard-core model. For general graphs, the best known bound is that achieved by the original Dobrushin condition, which establishes uniqueness for $\lambda < \frac{1}{b}$. For the special case in which the underlying graph is the square integer lattice $\mathbb{Z}^2$ (which is, of course, of subexponential growth), the best known bound [KRS89] is that the Gibbs measure is unique for $\lambda < 1.185$. The proof in [KRS89] is computer-assisted and uses the Dobrushin-Shlosman condition, i.e., a special case of Theorem 3.5 above, where updates are heat-bath of $L \times L$ squares for some L. When the underlying graph is a regular tree (obviously not of subexponential growth), the uniqueness regime is completely known. Specifically, on a regular tree, the Gibbs measure is unique if and only if $\lambda \leq \frac{b^b}{(b-1)^{b+1}}$.

We start our analysis with a general discussion of how to convert an analysis of the type carried out in [DG00] to our setting. The analysis in [DG00] is based on path coupling, where a coupling of an update is given for every pair of current configurations that differ at exactly one site, and it is shown that the average Hamming distance between the two resulting configurations is strictly less than 1, i.e., the distance decreases. Translating to our notation, using Hamming distance is equivalent to setting $u_x = 1$ for every x, and the fact that the distance decreases in every step is equivalent to $I_{\leftarrow y} - w_{B(y)} < 0$ for every site y, or equivalently, $I_{\leftarrow y} \,/\, w_{B(y)} < 1$. Thus, this coupled update rule satisfies the condition in Theorem 3.5, except that the (uniform) collection of metrics is not summable. However, as we already explained in the remark following Theorem 3.5, if the update rule uses blocks of bounded diameter and if the underlying graph is of sub-exponential growth, then the uniform metric can be modified to be summable while still maintaining $I_{\leftarrow y} \,/\, w_{B(y)} < 1$ for every site y. (Recall that a graph is said to be of subexponential growth if the volume of balls in the graph grows subexponentially with their radius, or equivalently, if there exists a

vertex x_0 such that for every $\epsilon > 0$, $\sum_y (1+\epsilon)^{-|x_0-y|}$ is finite, where $|x_0 - y|$ stands for the graph distance between x_0 and y.)

From here onwards we repeat the description and analysis w.r.t. the uniform collection of metrics of the update rule given in [DG00], but we do it using our terminology. Recall that in the hard-core model a configuration specifies a subset of occupied sites. It is therefore useful to write $\sigma \cup \{x\}$ for the configuration in which the subset of occupied sites is as in σ, except that x is also occupied. Similarly, $\sigma \setminus \{x\}$ stands for the configuration in which x is not occupied. The collection of blocks we use for the update rule is as follows: there is a block Θ_z for every site z, and Θ_z consists of z and all neighbors of z. As usual, the weight of each block is set to 1. As before, since blocks are indexed by sites, we write κ_z^σ for the distribution of an update of Θ_z in current configuration σ. The result of an update of Θ_z depends on the current configuration on the neighbors of z. Specifically, κ_z^σ is defined as the distribution resulting from the following update of Θ_z:

- If all the neighbors of z are unoccupied under σ: with probability $\frac{\lambda}{1+\lambda}$ the resulting configuration is $\sigma \cup \{z\}$, and with probability $\frac{1}{1+\lambda}$ the resulting configuration is $\sigma \setminus \{z\}$.
- If at least two of the neighbors of z are occupied under σ: the resulting configuration is deterministically set to $\sigma \setminus \{z\}$.
- If exactly one neighbor of z is occupied under σ, say this is x: with probability $\frac{\lambda}{4(1+\lambda)}$ the resulting configuration is $(\sigma \setminus \{x\}) \cup \{z\}$, and with probability $1 - \frac{\lambda}{4(1+\lambda)}$, the resulting configuration is $\sigma \setminus \{z\}$.

Notice that κ_z^σ is defined for all σ, not just feasible ones. (In [DG00], the update was defined only for feasible current configurations σ[3].) It is easy to verify that $\mu^\tau_{\Theta_z}$ is stationary w.r.t. κ_z for every feasible τ (since κ_z is reversible w.r.t. $\mu^\tau_{\Theta_z}$ for every feasible τ).

From the definition above it is easy to see that κ_z^σ depends neither on the spin of z itself nor on the configuration on $\partial\Theta_z$, i.e., it depends only on the configuration of the neighbors of z. With that in mind, we go on to define the coupling $K_z(\sigma, \xi)$ for pairs (σ, ξ) the differ only at y, where $y \in \Theta_z \cup \partial\Theta_z$. Since κ_z^σ does not depend on σ_z or on $\sigma_{\partial\Theta_z}$, in case $y = z$ or $y \in \partial\Theta_z$ we define $K_z(\sigma, \xi)$ as the coupling in which the two configurations

[3]Strictly speaking, for the hard-core model, it is possible to slightly modify our construction of the path coupling in Section 3.2.2 so that it would be enough to define the update rule (and the couplings K_z) only for feasible configurations. Nevertheless, we define the update rule for any current configuration σ so that we can use the general form of our theorems.

agree on Θ_z with probability 1. When $y \in \Theta_z$ and $y \neq z$, i.e., y is a neighbor of z, $K_z(\sigma, \xi)$ is defined as follows. Recall that σ and ξ agree on all sites other than y and assume w.l.o.g. that y is occupied under σ and unoccupied under ξ. $K_z(\sigma, \xi)$ is then defined according to the number of neighbors of z other than y that are occupied:

- If two or more of the neighbors of z other than y are occupied then both κ_z^σ and κ_z^ξ are deterministic, so there is a unique coupling of these two distributions. This is the coupling in which, with probability one, the resulting pair of configurations is $(\sigma \setminus \{z\}, \xi \setminus \{z\})$. Notice that in this case $\rho_{\Theta_z}(K_z(\sigma, \xi)) = 1$.

- If exactly one neighbor other than y is occupied (say, x is the occupied neighbor) then κ_z^σ is still deterministic so there is a unique coupling of κ_z^σ and κ_z^ξ, the one in which with probability $1 - \frac{\lambda}{4(1+\lambda)}$ the resulting pair of configurations is $(\sigma \setminus \{z\}, \xi \setminus \{z\})$, and with probability $\frac{\lambda}{4(1+\lambda)}$ the resulting pair is $(\sigma \setminus \{z\}, (\xi \setminus \{x\}) \cup \{z\})$. Notice that in this case $\rho_{\Theta_z}(K_z(\sigma, \xi)) = 1 - \frac{\lambda}{4(1+\lambda)} + 3\frac{\lambda}{4(1+\lambda)} = 1 + \frac{\lambda}{2(1+\lambda)}$.

- If all neighbors of z other than y are unoccupied then $K_z(\sigma, \xi)$ is the following coupling. With probability $\frac{1}{1+\lambda}$ the resulting pair is $(\sigma \setminus \{z\}, \xi \setminus \{z\})$, with probability $\frac{3\lambda}{4(1+\lambda)}$ the resulting pair is $(\sigma \setminus \{z\}, \xi \cup \{z\})$, and with probability $\frac{\lambda}{4(1+\lambda)}$ the resulting pair is $((\sigma \setminus \{y\}) \cup \{z\}, \xi \cup \{z\})$. Notice that $(\sigma \setminus \{y\}) \cup \{z\} = \xi \cup \{z\}$ and hence $\rho_{\Theta_z}(K_z(\sigma, \xi)) = \frac{1}{1+\lambda} + 2\frac{3\lambda}{4(1+\lambda)} = 1 + \frac{\lambda}{2(1+\lambda)}$ in this case.

We conclude that $\rho_{\Theta_z}(K_z(\sigma, \xi)) \leq 1 + \frac{\lambda}{2(1+\lambda)}$ if y is a neighbor of z, and $\rho_{\Theta_z}(K_z(\sigma, \xi)) = 0$ otherwise. In particular, $I_{\leftarrow y} = \deg(y)(1 + \frac{\lambda}{2(1+\lambda)})$.

Now, since y is included in $\deg(y) + 1$ blocks (which are Θ_y, and Θ_z for every neighbor z of y), $w_{B(y)} = \deg(y) + 1$. Thus, $I_{\leftarrow y} / w_{B(y)} = \deg(y)(1 + \frac{\lambda}{2(1+\lambda)})/(\deg(y) + 1)$, and $\sup_y I_{\leftarrow y} / w_{B(y)} < 1$ if $\frac{\deg(y)\lambda}{2(1+\lambda)} < 1$ for every y, i.e., if $\lambda < \frac{2}{b-1}$. Hence, using Theorem 3.5 and the explanation at the beginning of this subsection, the hard-core model with activity parameter λ on graphs of subexponential growth of maximum degree $b + 1$ admits a unique Gibbs measure for $\lambda < \frac{2}{b-1}$.

Remark: Notice that for the update rule described in this subsection, $I_{\leftarrow y} < \sum_x I_{x \leftarrow y}$. In other words, it was crucial for our analysis that in the definition of $I_{\leftarrow y}$, the quantification over configurations is taken only once before summing over x, rather than quantifying separately for each x. To see this, recall our analysis of the distance $\rho_{\Theta_z}(K_z(\sigma, \xi))$, and notice that for every pair of configurations (σ, ξ), there can be at most one site x other than z and y for which $\rho_x(K_z(\sigma, \xi)) > 0$

(specifically, this can only happen if x is the unique neighbor of z other than y that is occupied). However, when calculating $\sum_x I_{x \leftarrow y}$ we need to consider the worst pair of configurations for each x separately, and hence the coupling K_z contributes to the distance at all neighbors x of z. The last observation follows from the fact that for each x, we have to consider the pair of configurations in which x is the unique occupied neighbor of z other than y. When one considers the total influence on a site w.r.t. the above update rule, a similar issue arises. In particular, $I_{y \leftarrow} = \sum_x I_{y \leftarrow x} > I_{\leftarrow y}$, so we cannot use the same update rule in order to establish uniqueness of the Gibbs measure using Theorem 3.3 (which would apply to any underlying graph) for the same range of parameters. In fact, w.r.t. the above update rule, the condition based on the total influence on a site holds for an even smaller range of parameter values then the single-site Dobrushin condition.

Chapter 4

A combinatorial view of mixing on the integer lattice

In this chapter we discuss relationships between mixing in time and in space for systems on the integer lattice $\mathbb{Z}^d$. The "sub-exponential growth" of this lattice gives rise to a sharp equivalence between optimal temporal mixing of the Glauber dynamics (as in Definition 2.7), and strong spatial mixing (as in Definition 2.4). Notice that both forms of mixing in this equivalence are uniform in the boundary condition. (For boundary-specific equivalences, see Chapters 5 (for systems on trees) and 6 (for systems on the square integer lattice).)

Variants of the above equivalence have been explored by a number of previous authors, using various notions of mixing in both time and space. This line of work was initiated by Holley [Hol85] and Aizenman and Holley [AH87], followed by Zegarlinski [Zeg90] and culminating in the work of Stroock and Zegarlinski [SZ92], who were the first to establish the above equivalence in full. We further mention Martinelli and Olivieri [MO94a, MO94b], who later obtained sharper results by working with a weaker spatial mixing assumption, and Cesi [Ces01], who recently simplified some of the proofs. See also [Mar98] for a review of related results.

The references mentioned above make crucial use of functional analysis in their proofs, and usually discuss the quantities c_{gap} and c_{sob} (defined in Section 2.3.2) as a measure of mixing in time. In this chapter, we give purely combinatorial proofs of this equivalence, based on the elementary technique of coupling probability distributions. Although

some of the ideas we use have appeared before, our main contribution lies in presenting a complete argument which is purely combinatorial, where the reader does not need to resort to concepts from functional analysis.

We note that the result we present in the direction going from strong spatial mixing to optimal temporal mixing of the Glauber dynamics is limited in the sense that it only applies to *monotone* systems. For general systems, however, we show that strong spatial mixing implies optimal temporal mixing of the dynamics based on blocks of large enough radius. The corresponding implication for the (single-site) Glauber dynamics in the general case is known [Ces01, Mar98, MO94b, SZ92], but currently we do not have a combinatorial proof of it.

Before going on to the formal discussion, let us illustrate the above equivalence[1] by discussing it in the context of the Ising model (as defined in Example 2.1) on the integer lattice $\mathbb{Z}^d$. Recall that in the definition of the Ising model, β stands for the inverse temperature and h for an external field. The following fact is an example of the equivalence between temporal and spatial mixing: There exists a critical $\hat{\beta}_c$ such that, when $h = 0$ (no external field), for $\beta < \hat{\beta}_c$ both optimal temporal mixing uniformly in the boundary condition and strong spatial mixing hold, while for $\beta > \hat{\beta}_c$ both fail.

It is worth mentioning here that, in the special case of the Ising model on the square lattice $\mathbb{Z}^2$, the critical $\hat{\beta}_c$ mentioned above coincides [MOS94] with the critical inverse temperature β_c at which a phase transition occurs in the infinite volume limit, namely, for $\beta < \beta_c$ there exists a unique infinite volume Gibbs measure while for $\beta > \beta_c$ there are multiple such measures. Notice that in general it is *not* true that the two critical inverse temperatures $\hat{\beta}_c$ and β_c coincide, and there are examples where the infinite volume Gibbs measure is unique while strong spatial mixing does not hold (see [Mar98] for a discussion on this matter).

Still in the special case of the Ising model on $\mathbb{Z}^2$, the corresponding "phase transition" in the mixing time is known to be very sharp [CGMS96]. Specifically, for $\beta > \hat{\beta}_c = \beta_c$, not only does optimal temporal mixing not hold but in fact, for the free boundary condition, the mixing time is super-polynomial (specifically, $\Theta(m \exp(c\sqrt{n}))$ for some constant $c > 0$). For dimensions $d \geq 3$, a similar result (mixing time $\Theta(m \exp(cn^{1-1/d}))$ for the free

[1] Strictly speaking, the discussion in the following three paragraphs applies to slightly modified definitions of spatial and temporal mixing where the region Ψ is restricted to have a "nice" shape (see remark following Definition 2.4).

boundary) applies for low enough temperatures, but is not known to go all the way to the critical temperature, i.e., for all $\beta > \beta_c$.

The rest of the chapter is organized as follows. Section 4.1 includes definitions new to this chapter and precise statements of results. In Section 4.2 we list a few basic tools we use in the proofs. In Section 4.3 we prove that optimal temporal mixing implies spatial mixing, while in Section 4.4 we prove the converse, first for monotone systems and then for general systems.

4.1 Definitions and statements of results

Our focus in this chapter is on systems defined on the d-dimensional integer lattice[2], i.e., where the vertex set $V = \mathbb{Z}^d$ and $(v, u) \in E$, denoted $v \sim u$, if and only if $\sum_{i=1}^{d} |v_i - u_i| = 1$. The dynamics we focus on is Glauber (single site), but we consider other collections of blocks as well. Recall that we say that a collection of blocks $\{\Theta_i\}$ is of bounded diameter if there exists a constant r such that the diameter of any block Θ_i is at most r. We also discuss a specific collection of bounded-diameter blocks: let $Q_L = [1, \ldots, L]^d$ be the d-dimensional regular box of side length L; we write *HB*(L) for the (heat-bath) dynamics based on the collection of translations of Q_L, i.e., $\left\{\Theta_z = Q_L + z : z \in \mathbb{Z}^d\right\}$.

We denote instances of the dynamics by (X_t) (or (Y_t)), where X_t is the random variable denoting the configuration at time t, starting from X_0. We write $X_t[v]$ for the spin at v at time t, and similarly, $X_t[\Lambda]$ for the configuration on Λ at time t.

4.1.1 Monotone systems

Some of the statements in this paper apply only to monotone systems. In a monotone system, each site v is associated with a linear ordering of the spin space, denoted by $\succeq_v$. Since the spin space is finite, each of the linear orderings has unique maximal and minimal elements, which we call the $(+)$ and $(-)$ elements respectively. The single-site orderings give rise to a partial ordering $\succeq_\Psi$ of the configuration space. Specifically, $\tau^{(1)} \succeq_\Psi \tau^{(2)}$ if and only if $\tau_v^{(1)} \succeq_v \tau_v^{(2)}$ for every $v \in \Psi$. The system is *monotone* with respect to the above partial ordering if, for every region Ψ and any two boundary configurations $\tau^{(1)}$

[2]Most of our results hold — with suitable modifications — for any lattice with "sub-exponential growth" (i.e., the volume of increasing balls around any site increases subexponentially with the radius). For simplicity, in this chapter we focus just on $\mathbb{Z}^d$.

and $\tau^{(2)}$ such that $\tau^{(1)} \succeq_{\partial\Psi} \tau^{(2)}$, the Gibbs distribution $\mu_\Psi^{\tau^{(1)}}$ statistically dominates the Gibbs distribution $\mu_\Psi^{\tau^{(2)}}$ with respect to $\succeq_\Psi$. Equivalently, the two distributions can be coupled such that with probability 1, $\sigma^{(1)} \succeq_\Psi \sigma^{(2)}$, where $\sigma^{(1)}$ and $\sigma^{(2)}$ are a pair of coupled configurations chosen from $\mu_\Psi^{\tau^{(1)}}$ and $\mu_\Psi^{\tau^{(2)}}$ respectively. Notice that it is enough that the above property holds for all single sites to ensure that it holds for all regions Ψ. Also, since the single-site orderings are linear, the system is "realizably" monotone [FM01]. This means that, given a collection of boundary configurations $\tau^{(1)}, \tau^{(2)}, \ldots, \tau^{(k)}$, we can simultaneously couple the k corresponding Gibbs distributions such that if $\tau^{(i)} \succeq_{\partial\Psi} \tau^{(j)}$, the corresponding coupled configurations satisfy $\sigma^{(i)} \succeq_\Psi \sigma^{(j)}$ with probability 1 (simultaneously for each such pair i, j).

Many well known spin systems are monotone, including the Ising model and the hard-core model (independent sets).

4.1.2 Systems with hard constraints

Throughout this chapter, we assume the spin systems are permissive (see discussion in Section 2.1). In particular, when we say "arbitrary spin system", we mean "arbitrary permissive spin system". Recall that in permissive systems μ_Λ^τ is well defined for any choice of τ and Λ, and that the transitions of the Markov chains we discuss are well defined for any current configuration, even if it is not in the support of the stationary distribution. Furthermore, the chain is guaranteed to reach a configuration in the support at some finite time, and thus converge to the stationary Gibbs distribution. Hence, without loss of generality, we may think of the chains as running on the whole configuration space Ω_Ψ^η. In particular, when we say the dynamics has optimal temporal mixing (see Definition 2.7), we mean that the error bound applies to chains starting from an arbitrary configuration in Ω_Ψ^η, including those not in the support of μ_Ψ^η. Notice, however, that this has a negligible quantitative effect since once every site is updated at least once (which takes $O(m \log n)$ time with high probability) the configuration is guaranteed to be in the support of μ_Ψ^η.

4.1.3 Results

Several notions of mixing in time and in space for models on integer lattices are known to be equivalent to one another [Ces01, Mar98, MO94a, MO94b, SZ92], though the proofs are often rather complex and cast in the language of functional analysis. In this chapter we present combinatorial proofs of the following implications.

Theorem 4.1 *For any collection of bounded-diameter blocks $\{\Theta_i\}$, if the dynamics based on $\{\Theta_i\}$ has optimal temporal mixing for some boundary condition η then the system has weak spatial mixing; if in addition the optimal temporal mixing is uniform in the boundary condition then the system has strong spatial mixing.*

For monotone systems we show the converse as well:

Theorem 4.2 *If a monotone system has strong spatial mixing then the Glauber dynamics has optimal temporal mixing uniformly in the boundary condition.*

In the general case (without assuming monotonicity), we show:

Theorem 4.3 *If a system has strong spatial mixing then there exists a finite integer L for which the heat-bath block dynamics HB(L) has optimal temporal mixing uniformly in the boundary condition.*

Notice that strong spatial mixing implies optimal temporal mixing of the (single-site) Glauber dynamics in the general case as well [Ces01, Mar98, MO94b, SZ92], but we have not yet been able to find a purely combinatorial proof of this implication. The main obstacle is translating the rapid mixing result for the block dynamics into rapid mixing of the single-site dynamics (at the cost of only a constant factor), a problem which is still open for general spin systems. The functional analysis proofs mentioned above analyze the log-Sobolev constant, and show that strong spatial mixing implies bounded c_{sob} of *HB*(L), uniformly in the boundary condition; the implication for the single-site dynamics follows since c_{sob} of the latter is easily seen to differ from that of *HB*(L) by at most a constant factor (that depends on L).

4.2 Preliminaries

In this section we describe some of the common tools we use in our proofs.

4.2.1 Coupling and mixing time

As we saw in Chapter 3, a common tool for bounding the total variation distance between two distributions, and in particular for bounding the mixing time of Markov chains, is coupling. Recall that a coupling of ν_1 and ν_2 is any joint distribution whose marginals are ν_1

and ν_2 respectively. If σ_1 and σ_2 are a pair of random configurations chosen from a given coupling of ν_1 and ν_2 then $\Pr(\sigma_1 \neq \sigma_2)$ is an upper bound on the total variation distance between ν_1 and ν_2. Also, there is always an optimal coupling, i.e., a coupling such that $\Pr(\sigma_1 \neq \sigma_2) = \|\nu_1 - \nu_2\|$.

In the proofs we give in this chapter we use the following coupling of the dynamics, which we call an *identity coupling*. This coupling is similar to those used in Chapter 3, but here we use a version that allows us to simultaneously couple any number of instances of the chain. An identity coupling is determined by specifying, for each block Θ, a coupling of all the Gibbs distributions (ranging over all possible values for the configuration of $\partial\Theta$). Let $\left\{\tau^{(1)}, \ldots, \tau^{(k)}\right\} = \mathcal{S}^{\partial\Theta}$ denote the set of all possible configurations on $\partial\Theta$. An identity coupling is given by specifying a joint distribution κ_Θ whose marginals are $\mu_\Theta^{\tau^{(1)}}, \ldots, \mu_\Theta^{\tau^{(k)}}$. Given κ_Θ, we couple a collection of instances of the Glauber dynamics $(X_t^1), (X_t^2), \ldots, (X_t^\ell)$ using a Markovian coupling (i.e., the joint distribution of $X_{t+1}^1, \ldots, X_{t+1}^\ell$ is a function only of the coupled configurations $X_t^1, \ldots, X_t^\ell$) where the coupled transition $(X_t^1, \ldots, X_t^\ell) \rightarrow (X_{t+1}^1, \ldots, X_{t+1}^\ell)$ is as follows:

- Choose a block Θ u.a.r. from those that intersect Ψ (the same block for all chains).
- Choose a collection $(\sigma^{(1)}, \ldots, \sigma^{(k)})$ of configurations on Θ from the joint distribution κ_Θ.
- For every $1 \leq i \leq \ell$ set $X_{t+1}^i[\Theta] = \sigma^{(j)}$ if and only if $X_t^i[\partial\Theta] = \tau^{(j)}$.

An important property of this coupling is that, if $X_t^i[\partial\Theta] = X_t^j[\partial\Theta]$, then $X_{t+1}^i[\Theta] = X_{t+1}^j[\Theta]$ with probability 1. Notice that in a monotone system there exists a *monotone identity coupling*, i.e., a joint distribution κ_Θ such that whenever $\tau^{(i)} \succeq_{\partial\Theta} \tau^{(j)}$, $\sigma^{(i)} \succeq_\Theta \sigma^{(j)}$ with probability 1.

We say that an identity coupling has *optimal mixing* for a boundary condition η if for any region Ψ and any two instances $(X_t), (Y_t)$ of the dynamics for μ_Ψ^η, we have $\Pr(X_{km} \neq Y_{km}) \leq Cn\exp(-\alpha k)$ for some constants C and $\alpha > 0$, where the probability space is the coupling of X_{km} and Y_{km} resulting from the identity coupling of the two processes. Notice that optimal mixing of an identity coupling implies optimal temporal mixing of the dynamics. Finally, the *coupling time* of an identity coupling is the minimum T such that $\Pr(X_T \neq Y_T) \leq \frac{1}{e}$ for every pair of instances $(X_t), (Y_t)$. As a result, $\Pr(X_{kT} \neq Y_{kT}) \leq e^{-k}$ for any positive integer k.

4.2.2 Bounding the speed of propagation of information

A central idea in the analysis of the mixing time of the Glauber dynamics, in particular when using spatial mixing assumptions, is to bound the speed at which information propagates during the dynamical process. In this section we give a lemma of this sort following an argument explained to us by van den Berg, based on the idea of *paths of disagreement* (also known as *disagreement percolation* [Ber93]). The idea of bounding the speed of propagation of information originally appeared in [SZ92], and similar bounds can also be found in [Mar98, KMP01]. The quantitative analysis in the argument in fact goes back to the Richardson model [Ric73]. Our version below applies to the Glauber dynamics on general graphs (as in [KMP01]), rather than just $\mathbb{Z}^d$.

Lemma 4.4 *Consider an arbitrary spin system on an arbitrary graph of maximum degree $b+1$. Let (X_t) and (Y_t) be two copies of the Glauber dynamics on an arbitrary region Ψ such that the two initial configurations agree everywhere except on some region A. Let B be another region and let $r = \mathrm{dist}(A,B)$. Then, for any positive integer $k \leq \frac{r}{be^2}$, if we run the dynamics for $T = km$ steps, $\Pr(X_T[B] \neq Y_T[B]) \leq 4\min\{|A|,|B|\}\,(\frac{bek}{r})^r$, where the probability space is the coupling of X_T and Y_T resulting from any identity coupling of (X_t) and (Y_t). In particular, if $T = km$ and $\mathrm{dist}(A,B) \geq be^2k$, then $\Pr(X_T[B] \neq Y_T[B]) \leq 4\min\{|A|,|B|\}\,e^{-\mathrm{dist}(A,B)}$.*

In words, Lemma 4.4 states that in km steps, with high probability, information percolates a distance of at most be^2k. As we discuss in more detail following the proof, a slight modification of the lemma applies to dynamics based on an arbitrary collection of bounded-diameter blocks, where w.h.p. information does not percolate a distance larger than linear in k, with the constant depending on the collection of blocks.

Proof: Since we couple X_t and Y_t using an identity coupling, if at time zero v had the same spin in both chains and at time T the spins at v differ then it must be the case that at some time $t' \leq T$ the site chosen to be updated was v and immediately before the update of v at time t' the two chains had different spins at one of the neighbors of v. Carrying this argument inductively, if we assume that at time zero the only sites whose spins may differ are included in A then in order for a site v to have different spins at time T there must be a *path of disagreement* going from A to v. Specifically, there must be $v_0, v_1, \ldots, v_\ell = v$ and $0 < t_1 < t_2 < \ldots < t_\ell \leq T$ such that $v_0 \in A$ and for $1 \leq i \leq \ell$, $v_i \sim v_{i-1}$ and at time t_i the site chosen to be updated was v_i. Notice that for a given path $v_0, \ldots, v_\ell$ the probability

of this event occurring is at most $\binom{T}{\ell}(\frac{1}{m})^\ell$. Now, if the two configurations at time T differ at some site in B, there must be a path of disagreement of length at least $r = \mathrm{dist}(A, B)$ going from A to B. Since the number of (simple) paths of length ℓ going from A to B is bounded from above by $\min\{|A|, |B|\}\,(b+1)b^{\ell-1}$, we can conclude that the probability of a disagreement in B at time $T = km$ is at most

$$\begin{aligned}\min\{|A|,|B|\}\cdot\frac{b+1}{b}\cdot\sum_{\ell=r}^{km} b^\ell\binom{km}{\ell}\left(\frac{1}{m}\right)^\ell &\leq \min\{|A|,|B|\}\cdot\frac{b+1}{b}\cdot\sum_{\ell=r}^{\infty}\left(\frac{bek}{\ell}\right)^\ell \\ &\leq 4\min\{|A|,|B|\}\left(\frac{bek}{r}\right)^r,\end{aligned}$$

where in the first inequality we use Sterling's approximation, and in the second we used the fact that $r \geq be^2k$. □

Remark: Notice that in the above proof we did not assume that X_t and Y_t have the same boundary configuration outside Ψ. Indeed, the region A of initially disagreeing spins need not be included in Ψ. Furthermore, we will often use Lemma 4.4 in a setting where some of the sites of Ψ are held fixed throughout the process. Notice that the proof above is still valid in this setting (regardless of whether or not the fixed spins disagree, i.e., are of sites in A). In fact, it is valid even if the two compared chains have different sets of fixed sites, as long as the sites which are fixed in only one of the chains are all included in the region A, i.e., we just assume that the spins of these sites disagree in the two chains. An important point to keep in mind in these scenarios is that in both chains the probability of a given site (or block) being updated in each step should still be $1/m$. The scenarios mentioned in this remark will become clearer when they arise in the proofs below.

We conclude this section by discussing the applicability of Lemma 4.4 to dynamics based on collections of blocks other than single sites. Notice that the arguments in the proof are still valid in this setting, but now, rather than a disagreement path of sites, we have to consider a disagreement path of blocks, where two blocks Θ_i and Θ_j are neighbors if and only if $\Theta_i \cap \partial\Theta_j \neq \emptyset$. (Notice that this is indeed a symmetric relation). Namely, if we construct a new graph where each vertex is a block and the edges are given according to the above relation, then it is not too difficult to see that the rate of percolation of disagreements in the new graph is bounded as in Lemma 4.4. Notice, however, that we have to consider the maximum degree of the block graph, which may be larger than that of the original graph. Similarly, the distance between A and B that we appeal to in the proof is the shortest distance in the block graph between two blocks that intersect A and B respectively, and this distance may be shorter than $\mathrm{dist}(A, B)$ in the original graph. Nevertheless, if the blocks

are of bounded diameter, then the maximum degree of the block graph is bounded, and the block distance between A and B is linear in the original distance between them. Thus it is still true that, in km steps, disagreements do not percolate a distance larger than ck for some constant c, with very high probability. (The probability they do is again exponentially small in this distance.)

4.3 From temporal to spatial mixing

In this section we prove Theorem 4.1, which states that if a dynamics based on a collection of bounded-diameter blocks has optimal temporal mixing then spatial mixing holds. Theorem 4.1 follows by combining the two theorems stated below. The first of these is specific to integer lattices (or to graphs of subexponential growth), and states that on such graphs and for a dynamics based on bounded-diameter blocks, if optimal temporal mixing holds then optimal *projected* temporal mixing (as in Definition 2.8) holds. The second theorem states that on any underlying graph, optimal projected temporal mixing of a dynamics as above implies strong spatial mixing.

Theorem 4.5 *Consider an arbitrary spin system on $\mathbb{Z}^d$, a collection of bounded-diameter blocks $\{\Theta_i\}$, and a (boundary) configuration η. If the dynamics based on $\{\Theta_i\}$ has optimal temporal mixing for η then it also has optimal projected temporal mixing for η.*

Theorem 4.6 *For any spin system on an any underlying graph, and for any collection of bounded-diameter blocks $\{\Theta_i\}$:*

(i) *if the dynamics based on $\{\Theta_i\}$ has optimal projected temporal mixing for* some *boundary condition η, then the system has weak spatial mixing;*

(ii) *if in addition optimal projected temporal mixing holds* uniformly *in the boundary condition, then the system has strong spatial mixing.*

Remark: We note that, as will be apparent from the proof, Theorem 4.5 holds for any graph of subexponential growth rather than just $\mathbb{Z}^d$. However, subexponential growth of the underlying graph *is* required for the theorem to hold. Indeed, this theorem breaks down on trees. Explicit counterexamples can be found in Chapter 5, where we discuss various settings on trees for which optimal temporal mixing holds but the Gibbs measure is not unique, and in particular weak spatial mixing and hence optimal projected temporal mixing do not hold.

The proof of Theorem 4.1 clearly follows by combining Theorems 4.5 and 4.6, and we proceed with the proofs of the latter two. In order to make for an easier reading, both proofs are given rigorously only for the Glauber dynamics; when appropriate, we make note of the modifications needed in order for the arguments to hold for a dynamics based on an arbitrary collection of bounded-diameter blocks.

Proof of Theorem 4.5: Recall that the optimal temporal mixing assumption means that there exist constants C and $\alpha > 0$ such that for every Ψ, and any instance (X_t) of the dynamics for μ_Ψ^η,

$$\|X_{km} - \mu_\Psi^\eta\| \leq C|\Psi| \exp(-\alpha k).$$

By the triangle inequality, we can assume w.l.o.g. that for any two instances (X_t) and (Y_t) of the dynamics,

$$\|X_{km} - Y_{km}\| \leq C|\Psi| \exp(-\alpha k).$$

We have to show that under this assumption, there exist constants C' and $\alpha' > 0$ such that for every Ψ and $\Lambda \subseteq \Psi$, and any instance (X_t) of the dynamics for μ_Ψ^η,

$$\|X_{km} - \mu_\Psi^\eta\|_\Lambda \leq C'|\Lambda| \exp(-\alpha' k).$$

We will in fact show that for any two instances (X_t) and (Y_t),

$$\|X_{km} - Y_{km}\|_\Lambda \leq C'|\Lambda| \exp(-\alpha' k).$$

The idea of the proof is one we use throughout this paper, which involves using Lemma 4.4 in order to localize the dynamics. Namely, when we run the dynamics for km steps, with high probability information from sites which are at distance at least $(2d-1)e^2k$ from Λ does not percolate into Λ. Therefore, if we take a region Λ_k surrounding Λ and whose boundaries are at distance at least $(2d-1)e^2k$ from Λ, we can assume that the sites on the boundary of Λ_k are fixed throughout the process. Thus, we can use the optimal temporal mixing bound for a dynamics on the local region Λ_k, whose volume is smaller than that of Ψ. As shown below, the fact that the volume of Λ_k grows only subexponentially in k (this is where where we use the subexponential growth of $\mathbb{Z}^d$) gives the required bound. An additional point we need to make in order to carry out the above argument is that, when running the dynamics on Ψ, with high probability an appropriate portion of the time is spent in the region Λ_k. This, however, is an easy consequence of the Chernoff bound.

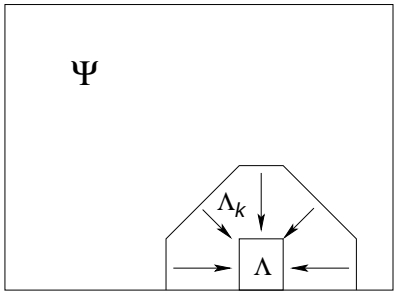

Figure 4.1: The region Λ_k, consisting of all sites in Ψ that are within distance $(2d-1)e^2k$ from Λ.

We proceed with the formal proof. Consider the region of all sites within distance $(2d-1)e^2k$ from Λ, and let Λ_k be the intersection of this region with Ψ (see Figure 4.1). Notice that $\mathrm{dist}(\Lambda, \Psi \setminus \Lambda_k) \geq (2d-1)e^2k$ and that $|\Lambda_k| \leq [2(2d-1)e^2k]^d|\Lambda|$.

Recall that we denote by m the number of blocks from which the dynamics chooses a block to update in each step, and that $m \equiv m(\Psi)$ (i.e., $m(\Psi)$ is the number of blocks intersecting Ψ). Since in the proof we consider processes both on Ψ and on Λ_k, we write $m[\Lambda_k]$ for $m(\Lambda_k)$ (and retain the notation m for $m(\Psi)$.)

In addition to the chains (X_t) and (Y_t), we consider two additional chains, denoted by $(X_t^{\Lambda_k})$ and $(Y_t^{\Lambda_k})$, whose initial configurations inside Λ_k are the same as (X_t) and (Y_t) respectively. The configuration outside Λ_k is given by η in both $(X_t^{\Lambda_k})$ and $(Y_t^{\Lambda_k})$ and remains fixed throughout the process, i.e., $(X_t^{\Lambda_k})$ and $(Y_t^{\Lambda_k})$ represent modified processes where, in a given step, if the chosen block to be updated is outside Λ_k then the configuration on that block remains unchanged, while if it intersects Λ_k then the intersection with Λ_k is updated as usual. Notice that this modified process is the same as running the dynamics for sampling from $\mu^{\eta}_{\Lambda_k}$, except that the probability of a block being chosen at a given step is $\frac{1}{m}$ instead of $\frac{1}{m[\Lambda_k]}$.

Using the triangle inequality, we have

$$\|X_{km} - Y_{km}\|_\Lambda \;\leq\; \|X_{km} - X_{km}^{\Lambda_k}\|_\Lambda \;+\; \|X_{km}^{\Lambda_k} - Y_{km}^{\Lambda_k}\|_\Lambda \;+\; \|Y_{km}^{\Lambda_k} - Y_{km}\|_\Lambda. \tag{4.1}$$

Lemma 4.4 (together with the remark following it) gives a bound on the first and third terms in the r.h.s. of (4.1). To see this, couple (X_t) and $(X_t^{\Lambda_k})$ using an identity coupling, i.e., an update of a block in (X_t) is coupled with an update of the same block in $(X_t^{\Lambda_k})$, where we note that an update of a block outside Λ_k in $(X_t^{\Lambda_k})$ means that the configuration remains the same. Notice that at time zero the two chains agree on Λ_k. (Clearly, both agree

outside Ψ throughout the process). Disagreement may percolate from $\Psi \setminus \Lambda_k$ into the bulk of Λ_k as we run the chains, but since $\mathrm{dist}(\Lambda, \Psi \setminus \Lambda_k) \geq (2d-1)e^2 k$, we can use Lemma 4.4 to deduce that $\|X_{km} - X_{km}^{\Lambda_k}\|_\Lambda \leq 4|\Lambda| e^{-(2d-1)e^2 k}$. (Here we used the bound Lemma 4.4 gives for the Glauber dynamics; see the end of this proof for the modifications required for a dynamics based on an arbitrary collection of bounded-diameter blocks.) By the same argument, $\|Y_{km} - Y_{km}^{\Lambda_k}\|_\Lambda \leq 4|\Lambda| e^{-(2d-1)e^2 k}$.

It remains to bound $\|X_{km}^{\Lambda_k} - Y_{km}^{\Lambda_k}\|_\Lambda$. Recall that both these chains have η fixed outside Λ_k so we can use the optimal temporal mixing assumption for a process on Λ_k. Notice that, when running the chain $X_t^{\Lambda_k}$ for km steps, $\frac{m[\Lambda_k]}{m} \cdot km = km[\Lambda_k]$ of the steps hit Λ_k on average. Using a Chernoff bound, with probability at least $1 - \exp(-\frac{km[\Lambda_k]}{8})$, the number of steps that hit Λ_k is at least $\frac{km[\Lambda_k]}{2}$. Thus, we can use the bound optimal temporal mixing gives for running the dynamics on Λ_k for $\frac{km[\Lambda_k]}{2}$ steps. Specifically,

$$\begin{aligned}
\|X_{km}^{\Lambda_k} - Y_{km}^{\Lambda_k}\|_\Lambda &\leq \|X_{km}^{\Lambda_k} - Y_{km}^{\Lambda_k}\|_{\Lambda_k} \\
&\leq C|\Lambda_k| \exp\left(-\alpha \cdot \frac{k}{2}\right) + \exp\left(-\frac{km[\Lambda_k]}{8}\right) \\
&\leq C[2(2d-1)e^2 k]^d |\Lambda| \exp\left(-\alpha \cdot \frac{k}{2}\right) + \exp\left(-\frac{k}{8}\right).
\end{aligned}$$

To conclude, by applying the above three bounds on the three terms on the r.h.s. of (4.1) respectively, we get:

$$\begin{aligned}
\|X_{km} - Y_{km}\|_\Lambda &\leq 8|\Lambda| e^{-(2d-1)e^2 k} + C[2(2d-1)e^2 k]^d |\Lambda| \exp\left(-\alpha \cdot \frac{k}{2}\right) + \exp\left(-\frac{k}{8}\right) \\
&\leq C'|\Lambda| \exp(-\alpha' \cdot k),
\end{aligned}$$

for appropriate constants C' and $\alpha' > 0$, as required.

We conclude the proof by mentioning the modifications needed in order for the proof to hold for a dynamics based on any collection of bounded-diameter blocks. The only place where we used the fact that the dynamics is Glauber is in applying the bound from Lemma 4.4. When we consider other collections of blocks, we need to set the radius of Λ_k accordingly, i.e., such that information has insufficient time to percolate from $\Psi \setminus \Lambda_k$ to Λ. However, as long as the collection of blocks is of bounded diameter, it is enough to take the radius linear in k in order for the probability of a disagreement percolating from outside Λ_k into Λ to be exponentially small in k. It is easy to see that the rest of the proof carries through in the same manner assuming these properties of Λ_k. □

Proof of Theorem 4.6: Both parts of the theorem follow from the following claim:

Claim 4.7 *For any boundary condition η, if the dynamics given in Theorem 4.6 has optimal projected temporal mixing for η then there exist constants C and $\alpha > 0$ such that, for every Ψ, $\Lambda \subseteq \Psi$, $\Delta \subseteq \partial\Psi$, and all σ that agree with η off Δ,*

$$\|\mu_\Psi^\eta - \mu_\Psi^\sigma\|_\Lambda \leq C|\Lambda| \exp(-\alpha \cdot \mathrm{dist}(\Lambda, \Delta)). \tag{4.2}$$

Part (ii) of Theorem 4.6 follows immediately from this claim because, under the hypothesis of part (ii) and assuming the claim, (4.2) holds for all η, i.e., the system has strong spatial mixing. For part (i) we use the triangle inequality to conclude that, under the hypothesis of this part, for every σ and τ,

$$\|\mu_\Psi^\sigma - \mu_\Psi^\tau\|_\Lambda \;\leq\; \|\mu_\Psi^\sigma - \mu_\Psi^\eta\|_\Lambda + \|\mu_\Psi^\eta - \mu_\Psi^\tau\|_\Lambda \;\leq\; 2C|\Lambda| \exp(-\alpha \cdot \mathrm{dist}(\Lambda, \partial\Psi)),$$

i.e., the system has weak spatial mixing. We thus go on to prove Claim 4.7.

The idea of the proof is that, when running the Glauber dynamics, the time needed in order for the projected distribution on Λ to be close to the stationary one is less than the time it takes for the disagreements on Δ to percolate into Λ. Formally, consider the following two instances of the Glauber dynamics on Ψ. The first, denoted by Z_t, is an instance with η as the boundary configuration while the second, denoted by Z'_t, is an instance with σ as the boundary configuration. The initial configuration of Ψ in both chains is chosen from the distribution μ_Ψ^σ. Notice that this is the stationary distribution of Z'_t and therefore $Z'_t = \mu_\Psi^\sigma$ for all t.

Using the triangle inequality, we have

$$\|\mu_\Psi^\eta - \mu_\Psi^\sigma\|_\Lambda \;=\; \|\mu_\Psi^\eta - Z'_t\|_\Lambda \;\leq\; \|\mu_\Psi^\eta - Z_t\|_\Lambda + \|Z_t - Z'_t\|_\Lambda.$$

By letting $t = \frac{\mathrm{dist}(\Delta,\Lambda)}{(2d-1)e^2} \cdot m$ we can make sure both terms are small. We bound the first term using the projected temporal mixing assumption. Namely, for $t = \frac{\mathrm{dist}(\Delta,\Lambda)}{(2d-1)e^2} \cdot m$ we have $\|\mu_\Psi^\eta - Z_t\|_\Lambda \leq C'|\Lambda| \exp(-\alpha' \cdot \frac{\mathrm{dist}(\Delta,\Lambda)}{(2d-1)e^2})$, where C' and $\alpha' > 0$ are the constants in the definition of optimal projected temporal mixing. We use Lemma 4.4 in order to bound the second term. Notice that Z_t and Z'_t have the same initial distribution on Ψ, and thus they can be coupled such that at time zero they have the same configuration on Ψ with probability 1. We continue to couple the two processes using an identity coupling. Disagreement may

percolate from Δ, but since $\mathrm{dist}(\Delta,\Lambda) = (2d-1)e^2\frac{t}{m}$ then by Lemma 4.4, $\|Z_t - Z'_t\|_\Lambda \leq 4|\Lambda|e^{-\mathrm{dist}(\Delta,\Lambda)}$. Putting the two bounds together we get

$$\begin{aligned}\|\mu^\eta_\Psi - \mu^\sigma_\Psi\|_\Lambda &\leq C'|\Lambda|\exp\left(-\frac{\alpha'}{(2d-1)e^2}\cdot\mathrm{dist}(\Delta,\Lambda)\right) + 4|\Lambda|e^{-\mathrm{dist}(\Delta,\Lambda)} \\ &\leq C|\Lambda|\exp(-\alpha\cdot\mathrm{dist}(\Delta,\Lambda)),\end{aligned}$$

for some constants C and $\alpha > 0$, as required.

We conclude the proof by mentioning the modifications needed in order for the proof to hold for a dynamics based on an arbitrary collection of bounded-diameter blocks. Again, the only place where we used the fact that the dynamics is Glauber is in applying the bound from Lemma 4.4. Indeed, when we consider other collections of blocks, we need to set the parameter t accordingly, i.e., such that information does not have enough time to percolate from Δ into Λ. However, as long as the collection of blocks is of bounded diameter, it is enough to take t linear in $\mathrm{dist}(\Delta,\Lambda)\cdot m$ in order for the probability of a disagreement percolating from Δ to Λ to be exponentially small in $\mathrm{dist}(\Delta,\Lambda)$. Therefore, the proof carries through for such collections of blocks. □

Remark: Notice that the arguments in this section were essentially boundary specific, and the only place where we used uniformity in the boundary condition is in applying Claim 4.7 for each η separately in order to establish (4.2) for every η (i.e., strong spatial mixing). Indeed, it is not too difficult to see from the same claim that optimal projected temporal mixing for η implies exponential decay of correlations for η (as in Definition 2.5). In fact, it is known [KMP01] that a weaker condition is enough for the latter to hold: if c_{gap} of a dynamics based on bounded-diameter blocks is bounded for η then the Gibbs distribution conditioned on η exhibits exponential decay of correlations.

4.4 From spatial to temporal mixing

In this section we prove that strong spatial mixing implies optimal temporal mixing. We first prove this for monotone systems where optimal temporal mixing is established for the Glauber dynamics, and then for general systems, where optimal temporal mixing is established for the block dynamics *HB*(L) with large enough L.

4.4.1 The monotone case

In this section we show that in monotone systems, strong spatial mixing implies optimal temporal mixing of the Glauber dynamics uniformly in the boundary condition (Theo-

rem 4.2). We state two theorems whose combination gives Theorem 4.2. The first theorem, whose proof uses ideas from the proof of Theorem 4.2 of [MO94a], states that the strong spatial mixing assumption implies $O(m \log^2 n)$ coupling time of any monotone identity coupling, uniformly in the volume n and in the boundary configuration. The second theorem, which is based on Theorem 3.12 of [Mar98], states (for general systems) that if there exists n_0 for which the coupling time of any identity coupling of the Glauber dynamics on regions of volume n_0 is at most $m\frac{c}{\log n_0} n_0^{1/d}$ for an appropriate constant c, then this identity coupling has optimal mixing. In particular, any upper bound of $o(m\frac{n^{1/d}}{\log n})$ on the asymptotic coupling time immediately implies that the identity coupling has optimal mixing. Formally, the two theorems read as follows:

Theorem 4.8 *Strong spatial mixing implies that there exists a constant c such that the coupling time of any monotone identity coupling of the Glauber dynamics for any region Ψ and arbitrary boundary condition is at most $T(n) = cm(\log n)^2$, where $n = |\Psi|$.*

Theorem 4.9 *Fix a boundary condition η, and suppose there exists an identity coupling such that for all regions Λ of volume at most n_0, where n_0 is a sufficiently large constant, the coupling time of the dynamics for μ_Λ^η is at most $m\frac{1}{8(2d-1)e^2}\frac{n_0^{1/d}}{\log n_0}$. Then for all n, all regions Ψ of volume n, any two instances $(X_t), (Y_t)$ of the dynamics for μ_Ψ^η, and every integer k, $\Pr(X_{km} \neq Y_{km}) \leq |\Psi| \exp(-ck)$, where $c = 2(2d-1)e^2 n_0^{-\frac{1}{d}}$ and the probability space is the coupling of (X_t) and (Y_t) resulting from the identity coupling above; namely, this identity coupling has optimal mixing for η.*

Theorem 4.2 clearly follows from the combination of Theorems 4.8 and 4.9, and we proceed with the proofs of the latter two.

Proof of Theorem 4.8: As in our earlier arguments, the idea of the proof is again to localize the dynamics, which allows us to use inductive bounds from smaller regions. However, here we use strong spatial mixing to achieve the localization, rather than the bound on the speed of propagation of information from Lemma 4.4.

Fix a large enough n_0 (to be determined later). By choosing an appropriate constant $c = c(n_0)$, the coupling time statement is true for all $n \leq n_0$. This is a consequence of the fact that any two instances of the chain will coalesce in finite time under any monotone coupling, e.g., because eventually both instances will simultaneously reach a maximal or minimal state. We go on to show the statement of the theorem is valid for $n > n_0$, by

inductively assuming its validity for volumes $k \leq [2\alpha^{-1}\log(3eC(q-1)n)]^d$, where α, C are the constants in the definition of strong spatial mixing (Definition 2.4), and $q = |\mathcal{S}|$ is the size of the spin space.

Consider an arbitrary region Ψ of volume n and an arbitrary boundary condition η. Let (X_t) and (Y_t) be two instances of the Glauber dynamics for μ_Ψ^η with arbitrary initial configurations inside Ψ. We will show that after $T(n)$ steps, for every site $v \in \Psi$, the probability (under the monotone identity coupling) that the two spins at v differ is at most $\frac{1}{en}$, and therefore, the probability that the two configurations (on the whole of Ψ) differ is at most $\frac{1}{e}$, as required.

Consider the regular box of radius $\alpha^{-1}\log(3eC(q-1)n)$ around v, and let Λ_v be the intersection of this box with Ψ. Notice that if v is close to the boundary of Λ_v then it is also close to the boundary of Ψ, but in any case, $\mathrm{dist}(v, \partial\Lambda_v \setminus \partial\Psi) \geq \alpha^{-1}\log(3eC(q-1)n)$. This fact will be useful for applying the strong spatial mixing assumption later on. Let $k = |\Lambda_v|$ and notice that $k \leq [2\cdot\alpha^{-1}\log(3eCn)]^d$. We now introduce four additional chains that may only update sites in Λ_v. We will couple these chains along with (X_t) and (Y_t) such that, whenever the site chosen to be updated is outside Λ_v only X_t and Y_t are updated while the additional four chains remain unchanged. On the other hand, when the site to be updated belongs to Λ_v all six chains are updated simultaneously according to the monotone identity coupling. Below we describe the additional four chains and their initial configurations. Let $\eta^{\Psi,+}$ and $\eta^{\Psi,-}$ denote the configurations in which the spins of sites in Ψ are all-$(+)$ and all-$(-)$ respectively, and where the configuration outside Ψ is given by η. The four chains are:

1. Q_t^+ : the chain starting from $\eta^{\Psi,+}$.
2. Q_t^- : the chain starting from $\eta^{\Psi,-}$.
3. Z_t^+ : the chain whose initial configuration is chosen at random from the Gibbs distribution $\mu_{\Lambda_v}^{\eta^{\Psi,+}}$.
4. Z_t^- : the chain whose initial configuration is chosen at random from the Gibbs distribution $\mu_{\Lambda_v}^{\eta^{\Psi,-}}$.

Notice that we can simultaneously couple the six chains such that at time zero, with probability one, $Q_0^+ \succeq X_0 \succeq Q_0^-$, $Q_0^+ \succeq Y_0 \succeq Q_0^-$, and $Z_t^+ \succeq Z_t^-$. Since we use a

monotone identity coupling, we have by induction that these relations hold for all t. Thus, we have

$$\Pr(X_t[v] \neq Y_t[v]) \;\leq\; \Pr(Q_t^+[v] \neq Q_t^-[v]) \;\leq$$

$$\Pr(Q_t^+[v] \neq Z_t^+[v]) \;+\; \Pr(Z_t^+[v] \neq Z_t^-[v]) \;+\; \Pr(Z_t^-[v] \neq Q_t^-[v]).$$

We use the strong spatial mixing assumption to bound the middle term of the last expression. Notice that Z_t^+ and Z_t^- are stationary chains, and in particular, they are the Gibbs distributions $\mu_{\Lambda_v}^{\eta^{\Psi,+}}$ and $\mu_{\Lambda_v}^{\eta^{\Psi,-}}$ respectively. Thus, by the strong spatial mixing assumption,

$$\|Z_t^+ - Z_t^-\|_{\{v\}} \;=\; \|\mu_{\Lambda_v}^{\eta^{\Psi,+}} - \mu_{\Lambda_v}^{\eta^{\Psi,-}}\|_{\{v\}} \;\leq\; C\exp(-\alpha \cdot \mathrm{dist}(v, \partial\Lambda_v \setminus \partial\Psi)),$$

where we used the fact that $\eta^{\Psi,+}$ and $\eta^{\Psi,-}$ agree on $\partial\Psi$. This bound on the total variation distance does not guarantee the same bound on disagreement under the coupling because the coupling we use is not necessarily the optimal one. However, monotonicity guarantees that our coupling is within a factor of $q-1$ (recall that q is the size of the spin space) from the optimal coupling, as explained next. We embed the ordering associated with v in the linear ordering $1, 2, \ldots, q$ with integer arithmetic. Since the spins at v are coupled such that with probability one $Z_t^+[v] \geq Z_t^-[v]$, we have

$$\begin{aligned}
\Pr(Z_t^+[v] \neq Z_t^-[v]) &\leq \mathrm{E}(Z_t^+[v] - Z_t^-[v]) \\
&= \mathrm{E}(Z_t^+[v]) - \mathrm{E}(Z_t^-[v]) \\
&\leq (q-1)\|Z_t^+ - Z_t^-\|_{\{v\}} \\
&\leq (q-1)C\exp(-\alpha \cdot \mathrm{dist}(v, \partial\Lambda_v \setminus \partial\Psi)) \\
&\leq \frac{1}{3en}.
\end{aligned}$$

Notice that in the second inequality above we used an optimal coupling of $Z_t^+[v]$ and $Z_t^-[v]$ together with the fact that the oscillation of any function whose range is $[1, q]$ is at most $q-1$.

In order to complete the proof we have to show that $\Pr(Q_t^+[v] \neq Z_t^+[v]) \leq \frac{1}{3en}$ when $t = T(n)$ (by symmetry, the same will hold for the $(-)$-chains). Recall that both (Q_t^+) and (Z_t^+) are instances of chains in which the configuration outside Λ_v is fixed as $\eta^{\Psi,+}$ throughout the process, i.e., they are the same as instances of the dynamics for sampling from $\mu_{\Lambda_v}^{\eta^{\Psi,+}}$, except that the probability of a site being updated is $\frac{1}{m} = \frac{1}{|\Psi|}$ rather than $\frac{1}{m[\Lambda_v]} = \frac{1}{|\Lambda_v|}$. Notice that when running the identity coupling for $T(n) = cm(\log n)^2$ steps,

$\frac{m[\Lambda_v]}{m} cm(\log n)^2 = cm[\Lambda_v](\log n)^2$ steps hit Λ_v on average, and by a Chernoff bound, with probability at least $1 - \frac{1}{6en}$ the number of steps that hit Λ_v is at least

$$\frac{1}{2} cm[\Lambda_v](\log n)^2 = (2\log n) cm[\Lambda_v] \frac{\log n}{4} \geq (2\log n) cm[\Lambda_v](\log k)^2$$

for large enough n. If we assume that indeed Λ_v is hit this often then we can use the induction hypothesis, which gives a bound on the probability of a disagreement under the identity coupling of two instances of the dynamics for $\mu_{\Lambda_v}^{\eta^{\Psi,+}}$. Indeed, after $T(k) = cm[\Lambda_v](\log k)^2$ steps in Λ_v, the configurations (on the whole of Λ_v) disagree with probability at most $\frac{1}{e}$, and thus after $(2\log n)T(k)$ steps, they disagree with probability at most $\frac{1}{n^2}$. Hence, $\Pr(Q^+_{T(n)}[v] \neq Z^+_{T(n)}[v]) \leq \frac{1}{6en} + \frac{1}{n^2} \leq \frac{1}{3en}$ for large enough n, as required. □

Remark: The reader may have noticed that, by carrying through a more careful analysis in the above proof, one can get a slightly better bound — for example, $O(m \log n (\log\log n)^2)$ — on the coupling time. However, since in any case we reduce the coupling time to $O(m \log n)$ using Theorem 4.9, we chose to keep the calculations simpler by only showing a bound of $O(m \log^2 n)$.

Proof of Theorem 4.9: Consider the Glauber dynamics on Ψ with boundary condition η. We will show that for any two instances of the chain (X_t) and (Y_t) and any $v \in \Psi$ we have $\Pr(X_{km}[v] \neq Y_{km}[v]) \leq \exp(-ck)$ under the given identity coupling. Using a union bound, this implies that $\Pr(X_{km} \neq Y_{km}) \leq |\Psi| \exp(-ck)$.

Let $\ell_0 = \lceil \frac{1}{c} \rceil = \lceil \frac{n_0^{1/d}}{2(2d-1)e^2} \rceil$. As before, we will use Lemma 4.4 to localize the dynamics. Together with the hypothesis of the theorem, this will imply that after $\ell_0 m$ steps the spins at v agree with high probability. What we want, however, is that the probability of disagreement will continue to decay exponentially with the number of steps. Notice that such a result would follow if, once the spins at v agreed, they continued to agree through the rest of the process, but this is clearly not the case. However, using the subexponential growth of $\mathbb{Z}^d$ and another localization argument, we can show that if all the spins within a large enough radius around v agree at a given time, then the spins at v will continue to agree for sufficiently many steps (depending on the radius of agreement). Bootstrapping from the sufficiently small probability of disagreement after $\ell_0 m$ steps, we get the required exponential decay.

We proceed with the formal proof. Let $\rho(k) = \max_{X_0, Y_0, v \in \Psi} \Pr(X_{km}[v] \neq Y_{km}[v])$. We have the following two claims.

Claim 4.10 *Under the hypothesis of the theorem,*

$$\rho(\ell_0) \leq \frac{1}{e2^d(n_0+1)} \;=\; \frac{1}{e2^d([2(2d-1)e^2\ell_0]^d+1)}\,.$$

Claim 4.11 *Without any assumptions, for any k_1 and k_2,*

$$\rho(k_1+k_2) \leq [2(2d-1)e^2k_2]^d\rho(k_1)\rho(k_2) + 4e^{-k_2}.$$

Theorem 4.9 follows from the combination of the above two claims. To see this, let $\phi(k) = 2^d([2(2d-1)e^2k]^d+1)\cdot\max\left\{\rho(k), 2e^{-\frac{k}{2}}\right\}$. Using Claim 4.11, we have by an explicit calculation that $\phi(2k) \leq \phi(k)^2$. On the other hand, from Claim 4.10 we get that $\phi(\ell_0) \leq \frac{1}{e}$ (where we used the fact that ℓ_0 is large enough to handle the case of $\rho(\ell_0) < 2e^{-\frac{\ell_0}{2}}$). We then conclude that $\rho(k) \leq \phi(k) \leq \exp(-\frac{k}{\ell_0})$, as required. This completes the proof of Theorem 4.9 (assuming Claims 4.10 and 4.11). □

Proof of Claim 4.10: Let $v \in \Psi$ be any site. As in Theorem 4.5, the idea is to take a regular box of volume n_0 around v. Then, since we run the coupled chains for only $\ell_0 m$ steps, information from the boundary of this box does not have enough time to percolate to v. We can therefore assume the boundaries around this box are fixed. But then, the assumption of the theorem guarantees that the spins at v will agree with the required probability.

Formally, let Λ_v be the intersection of the regular box of volume n_0 around v with Ψ. Let $(X_t^{\Lambda_v})$ and $(Y_t^{\Lambda_v})$ be two chains whose initial configurations inside Λ_v agree with X_0 and Y_0 respectively, and whose configuration outside Λ_v is fixed to η throughout the process. We have $\Pr(X_t[v] \neq Y_t[v]) \leq \Pr(X_t[v] \neq X_t^{\Lambda_v}[v]) + \Pr(X_t^{\Lambda_v}[v] \neq Y_t^{\Lambda_v}[v]) + \Pr(Y_t^{\Lambda_v}[v] \neq Y_t[v])$. Notice that $\mathrm{dist}(v, \partial\Lambda_v \setminus \partial\Psi) \geq \frac{1}{2}n_0^{1/d} = (2d-1)e^2\ell_0$. Therefore, using Lemma 4.4, we have $\Pr(X_{\ell_0 m}[v] \neq X_{\ell_0 m}^{\Lambda_v}[v]) \leq 4e^{-(2d-1)e^2\ell_0}$.

We go on to bound $\Pr(X_{\ell_0 m}^{\Lambda_v}[v] \neq Y_{\ell_0 m}^{\Lambda_v}[v])$. Notice that since in both chains the configuration outside Λ_v is fixed to η, and since $|\Lambda_v| \leq n_0$, we can use the hypothesis of the theorem to bound the above probability. If we run the coupled chains for $\ell_0 m$ steps, then with probability at least $1 - \exp(-\frac{\ell_0}{8}m[\Lambda_v])$ the number of steps that hit Λ_v is at least $\frac{\ell_0}{2}m[\Lambda_v]$. If indeed that many steps hit Λ_v then according to the hypothesis of the theorem, $\Pr(X_t^{\Lambda_v}[v] \neq Y_t^{\Lambda_v}[v]) \leq e^{-2\log n_0} = n_0^{-2}$. Thus, $\Pr(X_{\ell_0 m}^{\Lambda_v}[v] \neq Y_{\ell_0 m}^{\Lambda_v}[v]) \leq n_0^{-2} + \exp(-\frac{\ell_0}{8})$. Putting this together with the result of the previous paragraph we get $\Pr(X_{\ell_0 m}[v] \neq Y_{\ell_0 m}[v]) \leq n_0^{-2} + \exp(-\frac{\ell_0}{8}) + 8e^{-(2d-1)e^2\ell_0} \leq \frac{1}{e2^d(n_0+1)}$ for sufficiently large n_0, as required. □

Proof of Claim 4.11: We use Lemma 4.4 once again, this time in the sense that in $k_2 m$ steps, information can percolate over a distance of at most $(2d-1)e^2 k_2$. Thus, if the spins of all the sites within that radius from v agree after $k_1 m$ steps, then the spin at v will continue to agree after $(k_1 + k_2)m$ steps with high probability.

Formally, let Λ_{v,k_2} be the intersection of the regular box of radius $(2d-1)e^2 k_2$ around v with Ψ, and let $\mathcal{A}$ stand for the event that $X_{k_1 m}[\Lambda_{v,k_2}] \neq Y_{k_1 m}[\Lambda_{v,k_2}]$. Then, using Lemma 4.4 we have

$$\Pr(X_{(k_1+k_2)m}[v] \neq Y_{(k_1+k_2)m}[v]) \leq (1 - \Pr(\mathcal{A}))4e^{-(2d-1)e^2 k_2} + \Pr(\mathcal{A})\rho(k_2).$$

The proof is concluded once we notice that $\Pr(\mathcal{A}) \leq |\Lambda_{v,k_2}|\rho(k_1) \leq [2(2d-1)e^2 k_2]^d \rho(k_1)$. □

Remark: Notice that, in fact, the proof of Theorem 4.9 gives the stronger property of optimal projected temporal mixing, as does Theorem 4.5. The hypothesis of Theorem 4.9 differs from that of Theorem 4.5 in two respects. On one hand, the hypothesis of Theorem 4.9 is stronger because it works with the coupling time of an identity coupling rather than with the mixing time in general. On the other hand, the hypothesis in Theorem 4.9 is weaker because the time bounds are weaker. The reason why a weaker time bound is sufficient for coupling time is that we can appeal to the union bound $\Pr(X_t[\Lambda] \neq Y_t[\Lambda]) \leq \sum_{v\in\Lambda} \Pr(X_t[v] \neq Y_t[v])$. We used this union bound twice, first when we reduced the proof to bounding the probability of disagreement at a single site, and second when we bounded the probability of the event $\mathcal{A}$. Notice that the corresponding inequality for the total variation distance is not necessarily true. Namely, we cannot in general assert that $\|X_t - Y_t\|_\Lambda \leq \sum_{v\in\Lambda} \|X_t - Y_t\|_{\{v\}}$. If this assertion were true then we could have done with assuming a fast mixing time (rather than a fast coupling time) and working with the total variation distance rather than with the probability of disagreement throughout the proof.

4.4.2 The general case

In this section we prove Theorem 4.3. Namely, we show that in general (without assuming monotonicity), strong spatial mixing implies that the heat-bath block dynamics has optimal temporal mixing if the blocks used are sufficiently large regular boxes. Using path coupling [BD97], the proof is reduced to showing that strong spatial mixing implies a strong version of the condition for uniqueness of the Gibbs measure given in Theorem 3.13. The fact that strong spatial mixing implies this condition was already proved in [DS85b], but we include a simple proof of it here.

Proof of Theorem 4.3: Consider the heat-bath dynamics *HB*(L) on a region Ψ of volume n with an arbitrary boundary configuration. Notice that L here is a large enough constant to be chosen later, and will depend only on the dimension d and the constants in the definition of strong spatial mixing. In particular, L is uniform in n and the boundary configuration. Using Theorem 3.13 (i.e., the path-coupling method [BD97]), it is enough to give couplings of the updates of the blocks, starting from two configurations σ, ξ that differ at a single site on the boundary of the block (these couplings were denoted $K_i(\sigma, \xi)$ in Chapter 3), and show that $\sup_\Psi \max_{y \in \Psi} \left\{ I^\Psi_{\leftarrow y} / L^d \right\} < 1$, where $I^\Psi_{\leftarrow y}$ is defined in Section 3.3.2, the metric according to which $I^\Psi_{\leftarrow y}$ is defined is Hamming distance, and where we used the fact that each site is included in exactly L^d blocks.

Recalling that $I^\Psi_{\leftarrow y}$ represents the total influence of y on all the blocks it is adjacent to, and since there are $2dL^{d-1}$ such blocks, it is enough to give a coupling $K_i(\sigma, \xi)$ of the update of $\Theta = \Theta_i \cap \Psi$ such that the resulting average Hamming distance in Θ is $< \frac{L}{2d}$. (Recall that Θ_i is a translation of a regular box of side length L.) Let u be the site at which σ and ξ differ, $r = \frac{1}{2}(\frac{L}{4d})^{\frac{1}{d}}$, $\Theta_r = \{v \in \Theta \,|\, \mathrm{dist}(v, u) \leq r\}$, and $\overline{\Theta_r} = \Theta \setminus \Theta_r$. By the strong spatial mixing assumption, $\|\mu^\sigma_\Theta - \mu^\xi_\Theta\|_{\overline{\Theta_r}} \leq C|\overline{\Theta_r}| \exp(-\alpha \cdot r) \leq L^{-d}$ for a large enough L. We can thus couple the update of Θ such that the two coupled configurations disagree on $\overline{\Theta_r}$ with probability at most L^{-d}. A trivial upper bound on the resulting average Hamming distance in Θ in this coupling is then $|\Theta_r| + L^{-d}|\overline{\Theta_r}| \leq \frac{L}{4d} + 1$. □

Remark: It is not too difficult to see that the calculation in the proof also establishes the dual condition based on total influence *on* a site that was given in Theorem 3.11.

Chapter 5

Boundary-specific mixing on trees

In this chapter we consider spin systems on trees, and develop a new general framework for proving optimal temporal mixing (i.e., $O(m \log n)$ mixing time) of the Glauber dynamics in this setting. (In fact, the framework is for proving the stronger property of bounded log-Sobolev constant.) The main technical ingredient is an implication stating that, if the correlation (in the Gibbs distribution) between the spin at the root of the tree and the configuration ℓ levels below goes to zero fast enough with ℓ (a form of spatial mixing condition), then c_{sob} is bounded. The main novelty of this implication is that it is sensitive to the boundary condition, and thus opens up for the first time the possibility of proving fast mixing times for *specific* boundary conditions in situations where the mixing time under other boundary conditions is known to be significantly slower. We note that the converse of the above implication holds as well, and we thus get a boundary-specific equivalence between appropriate forms of mixing in time and in space for systems on trees.

Our framework is further enhanced by a second ingredient, which gives a simple criterion for the above spatial mixing condition, and thus for bounded c_{sob}. This criterion requires that the product of two natural quantities, which we call κ and γ, is at most $\frac{1}{b}$, where b is the branching factor of the tree. The quantities κ and γ measure the rate at which a spin disagreement at one site (in two copies of the system) can percolate down and up the tree respectively. We are thus able to prove optimal temporal mixing in many interesting scenarios for which this was not previously known by simply calculating the values of κ and γ in these scenarios.

Our main motivating application is the Ising model on a regular tree with $(+)$-boundary condition on the leaves, for which we show that c_{sob} remains bounded (and

therefore the mixing time remains $O(m \log n)$) at *all* temperatures and external fields. (The mixing time on a tree is always polynomial in n, but with free boundary at low temperatures and zero external field the exponent grows arbitrarily large as $\beta \to \infty$.) We apply our framework to other models as well, and obtain optimal temporal mixing over a significantly wider range of parameter values than previously known for independent sets (and, more generally, for any two-spin system), colorings and the Potts model. This includes situations in which the mixing time is strongly dependent on the boundary condition, as well as situations in which fast mixing holds for all boundary conditions.

The rest of this chapter is organized as follows. In Section 5.1 we set a motivating context for our results by discussing the Ising model on trees and its properties as the temperature, external field and boundary condition vary. Our general framework is laid out in detail in Section 5.2. We then apply this framework in Section 5.3, where model-specific calculations of κ and γ are carried out.

5.1 The Ising model and boundary conditions

5.1.1 The Ising model on the integer lattice

Before giving the details of the Ising model on a regular tree, it is helpful to recall the high-level picture of the Ising model on $\mathbb{Z}^2$ as a reference point. As was already discussed in Chapter 4, for the Ising model on $\mathbb{Z}^2$ with no external field it is well known that a phase transition occurs at a certain known critical inverse temperature $\beta = \beta_c$: for $\beta < \beta_c$ (the "high temperature" region) there are no long-range correlations between spins and consequently there is a unique Gibbs measure independent of the boundary condition, while for $\beta > \beta_c$ (the "low temperature" region) correlations are present at arbitrary distances and there are two distinct Gibbs measures (or "phases"), corresponding to the $(+)$- and $(-)$-boundary conditions respectively. In fact, for $\beta < \beta_c$ strong spatial mixing holds, and therefore optimal temporal mixing holds uniformly in the boundary condition (see Chapter 4 for details), while for $\beta > \beta_c$ it is known that for the free boundary condition the mixing time of the Glauber dynamics on a square of volume n is $\exp(\Omega(\sqrt{n}))$ [CGMS96].

One of the most interesting questions left open by the above results is the *influence of the boundary condition* on the mixing time. The following has been conjectured [BM02, FH87]:

Conjecture 5.1 *In the presence of an all-$(+)$ boundary, the mixing time in $\mathbb{Z}^2$ (and in fact in $\mathbb{Z}^d$ for all d) should remain polynomial in n* at all temperatures.

This captures the intuition that the only obstacle to rapid mixing for $\beta > \beta_c$ is the long time required for the dynamics to get through the "bottleneck" between the $(+)$-phase and the $(-)$-phase; the presence of the $(+)$-boundary eliminates the $(-)$-phase and hence the bottleneck. Formalizing this intuition, however, has proved very elusive.

5.1.2 The Ising model on trees

As already mentioned, the main application of the framework developed in this chapter is a proof of a strong version of Conjecture 5.1, when the lattice $\mathbb{Z}^d$ is replaced by a regular tree. (In statistical physics, the tree is referred to as the "Bethe approximation" of the lattice). Specifically, we analyze the mixing time of the Glauber dynamics for the Ising model on a tree with $(+)$-boundary condition on its leaves, and show that it remains $O(m \log n)$ at all temperatures and external fields. (This contrasts sharply with the fact that with a free boundary, the mixing time at low temperatures (as $\beta \to \infty$) with no external field is $mn^{\Omega(\beta)}$ [KMP01].) This is apparently the first result that quantifies the effect of boundary conditions on the dynamics in an interesting scenario. We stress that, while the tree is simpler in some respects than $\mathbb{Z}^d$ due to the lack of cycles, in other respects it is more complex: e.g., it exhibits a "double phase transition" (see below). Moreover, the Ising model on trees has recently received a lot of attention as the canonical example of a statistical physics model on a "non-amenable" graph (i.e., one whose boundary is of comparable size to its volume) — see, e.g., [KMP01, BRSSZ01, BRZ95, EKPS00, Iof96a, JS99, Lyo00, ST98].

To better appreciate our results for the Ising model on regular trees, we now give a detailed overview of known facts about this model, regarding both the Gibbs state and the Glauber dynamics. Fix $b \geq 2$ and let $\mathbb{T}^b$ denote the infinite b-ary tree (in which every vertex has b children)[1]. Consider for now the Ising model with no external field on $\mathbb{T}^b$. This model is known to have *two* critical inverse temperatures, β_0 and β_1. The first of these, $\beta_0 = \frac{1}{2}\ln(\frac{b+1}{b-1})$, marks the dividing line between uniqueness and non-uniqueness of the Gibbs measure: i.e., the "high temperature" region, in which the Gibbs measure is unique, is defined by $\beta \leq \beta_0$ [Pre74]. However, in contrast to the model on $\mathbb{Z}^d$, there is now a

[1] In the Bethe lattice mentioned above all vertices (including the root) have degree $b+1$; for convenience we define $\mathbb{T}^b$ such that the root has degree b, and notice that all our results below apply equally well to the setting in which the root has $b+1$ children.

second critical point $\beta_1 = \frac{1}{2}\ln(\frac{\sqrt{b}+1}{\sqrt{b}-1})$ [BRZ95, Iof96a], which delimits the region where "typical" boundary conditions exert long-range influence on the root. I.e., there is now an "intermediate" region $\beta_0 < \beta \leq \beta_1$ in which the $(+)$- and $(-)$-boundaries exert long-range influence but typical boundaries do not, while in the "low temperature" region $\beta > \beta_1$ long-range influence occurs even for typical boundaries. β_1 has alternative interpretations as the critical value for extremality of the Gibbs measure and the threshold for noisy data transmission on the tree [EKPS00].

The Glauber dynamics for the Ising model on trees has also been studied. In a recent paper [KMP01], it is shown that the mixing time with free boundary on a complete b-ary tree with n vertices is $O(m \log n)$ at high and intermediate temperatures (i.e., when $\beta < \beta_1$)[2]. Moreover, as soon as $\beta > \beta_1$ the mixing time becomes $mn^{\Omega(\beta)}$, so that the exponent is unbounded as $\beta \to \infty$. Thus the critical point $\beta = \beta_1$ is reflected in a jump in the (scaled) mixing time from logarithmic to polynomial in n. We note that a closer look at the fast mixing result for high and intermediate temperatures in [KMP01] reveals that it in fact holds for all boundary conditions.

Now, when one considers the effect of boundary conditions, trees differ greatly from $\mathbb{Z}^d$ because their boundary is very large (of size $\Theta(n)$ rather than $\Theta(n^{1/d})$ as in $\mathbb{Z}^d$). However, this can be compensated for by introducing a non-zero external field h (which adds a bias to each internal spin in the direction of the field). It is well known [Geo88] that, for all $\beta > \beta_0$, there is a critical value $h = h_c(\beta) > 0$ of the field such that the Gibbs measure is not unique when $|h| \leq h_c$, and is unique when $|h| > h_c$ (see Fig. 5.1). (When $\beta \leq \beta_0$ the Gibbs measure is unique for all h, and h_c is defined to be zero.) Thus in the presence of a $(+)$-boundary, the tree with an external field of value $h = -h_c$ is the analog of the classical case of $\mathbb{Z}^d$ with zero field. (Notice that on $\mathbb{Z}^d$ the critical value h_c is always zero because the size of the boundary of a box is much smaller than its volume, and hence a non-zero external field overcomes the influence of any boundary condition).

Our result for the $(+)$-boundary condition (Theorem 5.2 below) applies to the full range of values of both β and h. The fact that we are able to handle external fields, and specifically the critical value $|h| = h_c$, brings our results for trees rather close to Conjecture 5.1 for $\mathbb{Z}^d$.

We conclude this section with the detailed results proven in this chapter for the

[2] Actually [KMP01] proves this only for sufficiently high temperatures, but the argument can be extended to all $\beta < \beta_1$ [PW02].

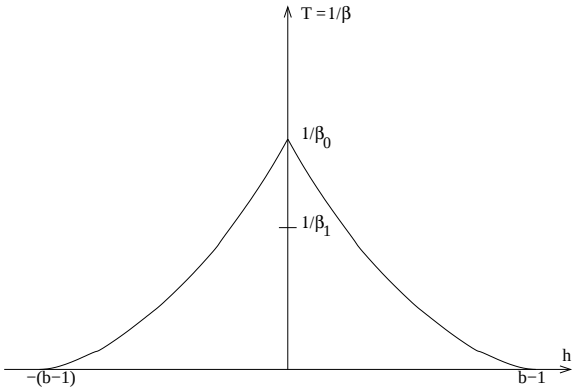

Figure 5.1: Curve of critical field $h_c(\beta)$. The Gibbs measure is unique above the curve.

Ising model on $\mathbb{T}^b$.

Theorem 5.2 *For any fixed b, the Glauber dynamics on the n-vertex b-ary tree with $(+)$-boundary condition has bounded c_{sob}, and therefore $O(m \log n)$ mixing time, at all inverse temperatures $\beta < \infty$ and all external fields h.*

In our second result for the Ising model, we obtain improved and more general bounds on the mixing time in cases where it is *insensitive* to the boundary condition, i.e., in the high and intermediate temperature region at all fields, and at all temperatures when there is a large external field:

Theorem 5.3 *For any fixed b, the Glauber dynamics on the n-vertex b-ary tree with arbitrary boundary conditions has bounded c_{sob}, and therefore $O(m \log n)$ mixing time, both (i) at all inverse temperatures $\beta < \beta_1$ and all external fields h; and (ii) at all inverse temperatures $\beta < \infty$ and all external fields $|h| > h_c(\beta)$.*

This result has a several advantages over previous ones [KMP01, PW02]: it applies also when there is an external field; it gives the stronger result of bounded c_{sob} (rather than just $O(m \log n)$ mixing time); and the proof goes through directly for all boundary conditions.

Finally, we note that the results in Theorems 5.2 and 5.3 establish examples where c_{sob} is bounded (and therefore optimal temporal mixing holds), but optimal projected temporal mixing does *not* hold. In Chapter 4 we saw (Theorem 4.6) that if optimal projected temporal mixing holds, even for a single boundary condition, then weak spatial mixing holds and therefore the Gibbs measure is unique. Observe that the results in Theorems 5.2

and 5.3 include regimes of parameters for which there are multiple Gibbs measures, and thus for these regimes optimal projected temporal mixing does not hold even though c_{sob} is bounded.

5.2 A general framework for optimal temporal mixing

In this section we present a simple criterion for bounded c_{sob} that applies to any permissive spin system on a regular tree. We also give a similar criterion for bounded c_{gap}. Although bounded c_{sob} implies bounded c_{gap}, we carry out the analysis for both because, on the one hand, most of the analysis is identical and allows for a unified treatment, while on the other hand, in those places where the analysis differs, the arguments for c_{gap} are more intuitive and help in understanding the corresponding arguments for c_{sob}. The main argument underlying the two criteria is an equivalence between two natural forms of mixing in space and bounded c_{gap} and c_{sob} respectively.

We now set up some definitions in order to state our criteria. Throughout this chapter, we consider the case in which the Glauber dynamics is run on a complete finite subtree T of $\mathbb{T}^b$, rooted at the root of $\mathbb{T}^b$; if T has depth k then it has $n = (b^{k+1}-1)/(b-1)$ vertices, and its boundary ∂T consists of the children (in $\mathbb{T}^b$) of its leaves, i.e., $|\partial T| = b^{k+1}$. Since the dynamics is always Glauber, we write $c_{\mathrm{gap}}(\mu_T^\eta)$ and $c_{\mathrm{sob}}(\mu_T^\eta)$ for $c_{\mathrm{gap}}(P)$ and $c_{\mathrm{sob}}(P)$ respectively, where P is the Markov kernel of the Glauber dynamics for μ_T^η. Also, when we say that c_{gap} is bounded for a boundary condition η, we mean that there exists $\alpha > 0$ such that $c_{\mathrm{gap}}(\mu_T^\eta) > \alpha$ for all complete subtrees T. (This differs slightly from the corresponding definition in Section 2.3.2, where the quantification is over all subsets.) A similar remark applies to bounded c_{sob} and to optimal temporal mixing.

The key ingredients in our criteria for bounded c_{gap} and c_{sob} are two quantities, which we call κ and γ, that bound the rate of percolation of disagreements down and up the tree respectively. Both are properties of the collection of Gibbs distributions $\{\mu_T^\eta\}$, where the boundary condition η is fixed and T ranges over all finite (complete) subtrees of $\mathbb{T}^b$. To define κ and γ we need a little notation. For a configuration $\tau \in \Omega_T^\eta$, let $\tau^{x,s}$ denote the configuration τ with the spin at x set to s. For a site $x \in T$, write T_x for the maximal subtree of T rooted at x. When x is not the root of T, let $\mu_{T_x}^s \equiv \mu_{T_x}^{\tau^{z,s}}$ denote the Gibbs distribution in which the parent z of x has its spin fixed to s. Notice that the configuration on the bottom boundary of T_x is in fact specified by η because $\tau \in \Omega_T^\eta$. Indeed, $\mu_{T_x}^s$ depends

on τ only in that the latter specifies the (fixed) configuration on $T \setminus (T_x \cup \{z\})$ in $\mu^s_{T_x}$, since the distribution inside T_x is independent of τ once the spin at z is fixed.

Definition 5.4 *For a collection of Gibbs distributions $\{\mu^\eta_T\}$ as above, define $\kappa \equiv \kappa(\{\mu^\eta_T\})$ and $\gamma \equiv \gamma(\{\mu^\eta_T\})$ by*

(i) $\kappa = \sup_T \max_{z,s_1,s_2} \|\mu^{s_1}_{T_z} - \mu^{s_2}_{T_z}\|_z$;

(ii) $\gamma = \sup_T \max \|\mu^{\tau^{y,s_1}}_A - \mu^{\tau^{y,s_2}}_A\|_z$, *where the maximum is taken over all subsets $A \subset T$, all boundary configurations τ, all sites $y \in \partial A$, all neighbors $z \in A$ of y, and all spins $s_1, s_2 \in S$.*

Remark: Note that κ is the same as γ, except that the maximization is restricted to $A = T_z$ and the boundary vertex y being the parent of z; hence always $\kappa \le \gamma$. Since κ involves Gibbs distributions only on maximal subtrees T_z, it may depend on the boundary condition η at the bottom of the tree. By contrast, γ bounds the worst-case probability of disagreement for an *arbitrary* subset A and arbitrary boundary configuration around A, and hence depends only on the potentials of the system and not on η. It is the dependence of κ on η that opens up the possibility of an analysis that is specific to the boundary condition. For example, for the Ising model at very low temperature and with no external field κ is close to 1 in the free boundary case, while it is close to zero in the $(+)$-boundary case.

The intuition for these definitions comes from the following claim, which relates κ and γ to the rate of disagreement percolation in the tree. For any T and site $x \in T$, write $\widetilde{T_x}$ for $T_x \setminus \{x\}$, the subtree T_x excluding its root, and $\mu^s_{\widetilde{T_x}} \equiv \mu^{\tau^{x,s}}_{\widetilde{T_x}}$ for the Gibbs distribution when the spin at x is fixed to s. Also, for $\ell \le \mathrm{height}(x) + 1$, write $B_{x,\ell}$ for the subtree (or "block") of height $\ell - 1$ rooted at x (i.e., $B_{x,\ell}$ has ℓ levels). For two configurations $\sigma, \sigma' \in \Omega$, let $|\sigma - \sigma'|_{x,\ell}$ denote the number of sites ℓ levels below x (i.e., on the bottom boundary of $B_{x,\ell}$) at which σ and σ' differ. Note that $|\sigma - \sigma'| \le b^\ell$.

Claim 5.5 *For every $x \in T$ and all $\ell \le \mathrm{height}(x) + 1$ the following hold*:

(i) *For all s_1, s_2, there is a coupling $\nu = \nu^{s_1,s_2}_x$ of $\mu^{s_1}_{\widetilde{T_x}}$ and $\mu^{s_2}_{\widetilde{T_x}}$ for which $\mathrm{E}_\nu|\sigma - \sigma'|_{x,\ell} \le (\kappa b)^\ell$, where (σ, σ') is a random pair of configurations chosen from the coupling ν.*

(ii) *For any $\tau, \tau' \in \Omega$ that have the same spin value at the parent of x, $\|\mu^\tau_{B_{x,\ell}} - \mu^{\tau'}_{B_{x,\ell}}\|_x \le \gamma^\ell \cdot |\tau - \tau'|_{x,\ell}$.*

The proof of this claim follows from a standard recursive coupling along paths in the tree and is given in Section 5.2.4. Part (i) shows that κ bounds the probability of a disagreement percolating *down* the tree: i.e., when we fix a disagreement at x and recursively couple the distributions on the children of x, the expected proportion of disagreements after ℓ levels is at most κ^ℓ. Similarly, from part (ii) we see that γ bounds the probability of a disagreement percolating *up* the tree: i.e., when we fix a single disagreement at level ℓ below x, the probability of this disagreement reaching x is at most γ^ℓ.

We now state our main theorem that gives criteria for bounded c_{gap} and c_{sob} in terms of the quantities κ and γ.

Theorem 5.6 *Consider an arbitrary (permissive) spin system and a boundary condition η (a configuration on $\mathbb{T}^b$). For $\kappa \equiv \kappa(\{\mu_T^\eta\})$ and $\gamma \equiv \gamma(\{\mu_T^\eta\})$ the following hold:*

(i) *If $\gamma\kappa b < 1$ then c_{gap} is bounded for η.*

(ii) *If $\max\{\gamma\kappa b, \gamma\} < 1$ then c_{sob} is bounded for η.*

The proof of Theorem 5.6, which is the content of the rest of this section, has two main parts. In the first, we define two natural forms of mixing in space involving variance and entropy respectively, and show that they imply (and are in fact equivalent to) bounded c_{gap} and bounded c_{sob} respectively. In this part, the arguments are almost identical for both quantities, and we carry out a unified analysis. In the second part of the proof, we use Claim 5.5 to show that the two forms of mixing in space hold when conditions (i) and (ii) respectively of Theorem 5.6 hold, thus concluding the proof of this theorem. In this part of the proof the analysis is carried out first for the spectral gap, where the arguments are much simpler, and then for the more involved log-Sobolev constant. At the end of the section we also show that Theorem 5.6 continues to hold when κ and γ are replaced by slightly modified (relaxed) versions of the two quantities. The relaxed version of the theorem is motivated by some of the applications we discuss in Section 5.3. Finally, we note that in the course of this section we occasionally cite specialized versions of theorems given in Appendix A, where various inequalities for variance and entropy are collected. For proofs of these theorems, the reader is referred to the general versions in Appendix A.

5.2.1 Relating mixing in time to mixing in space

In this section we show that two spatial mixing conditions are equivalent to bounded $c_{\rm gap}$ and bounded $c_{\rm sob}$ respectively. Our analysis has two main advantages over those used previously: first, the conditions for the spectral gap and the log-Sobolev constant are identical in form, allowing a uniform treatment; second, and more importantly, the conditions are *measure-specific*, i.e., they may hold for the Gibbs distribution induced by some specific boundary configuration while not holding for other boundary configurations. Hence, the conditions are sensitive enough to show rapid mixing for specific boundaries even though the mixing time with other boundaries is slow for the same choice of system potentials.

Reduction to block analysis

Before presenting the main result of this section, we need some more definitions and background. Recall that for a site $x \in T$, $B_{x,\ell} \subseteq T$ denotes the intersection with T of the subtree (or "block") of height $\ell - 1$ rooted at x, i.e., $B_{x,\ell}$ consists of ℓ levels. (Here we also consider the case in which x is $k < \ell$ levels from the bottom of T, where $B_{x,\ell}$ has only k levels.) In what follows we will think of ℓ as a suitably large constant. By analogy with expression (2.5) for the Dirichlet form, let $\mathcal{D}_\ell(f) \equiv \sum_{x \in T} \mu_T^\eta[\mathrm{Var}_{B_{x,\ell}}(f)]$ denote the local variation of f w.r.t. the blocks $\{B_{x,\ell}\}$. A straightforward manipulation (see, e.g., [Mar98], keeping in mind that each site belongs to at most ℓ blocks) shows that $c_{\rm gap}$ can be bounded as follows:

$$c_{\rm gap}(\mu_T^\eta) \geq \frac{1}{\ell} \cdot \inf_f \frac{\mathcal{D}_\ell(f)}{\mathrm{Var}_T^\eta(f)} \cdot \min_{\tau,x} c_{\rm gap}(\mu_{B_{x,\ell}}^\tau). \tag{5.1}$$

As before, the infimum is taken over non-constant functions (and henceforth we omit explicit mention of this). The importance of (5.1) is that $\min_{\tau,x} c_{\rm gap}(\mu_{B_{x,\ell}}^\tau)$ depends only on the size of $B_{x,\ell}$ and the potentials of the system, but not on the size of T; in fact, it is at least $\Omega(e^{-c\ell})$, where the constant c depends on the potentials [KMP01]. Therefore, in order to show that $c_{\rm gap}$ is bounded by a constant independent of the size of T, it is enough to show that, for some finite ℓ, $\mathrm{Var}_T^\eta(f) \leq \mathrm{const} \times \mathcal{D}_\ell(f)$ for all functions f. This is what we will show below, under the relevant spatial mixing condition. As a side remark, notice that $\mathcal{D}_\ell$ is simply the Dirichlet form of the heat-bath dynamics based on the blocks $B_{x,\ell}$, and hence $\inf_f \frac{\mathcal{D}_\ell(f)}{\mathrm{Var}_T^\eta(f)}$ is exactly the (scaled) spectral gap of this dynamics.

An identical manipulation yields an analogous bound for the log-Sobolev constant.

For a non-negative function f, let $\mathcal{E}_\ell(f) \equiv \sum_{x\in T} \mu_T^\eta[\mathrm{Ent}_{B_{x,\ell}}(f)]$. Then

$$c_{\mathrm{sob}}(\mu_T^\eta) \geq \frac{1}{\ell} \cdot \inf_{f\geq 0} \frac{\mathcal{E}_\ell(f)}{\mathrm{Ent}_T^\eta(f)} \cdot \min_{\tau,x} c_{\mathrm{sob}}(\mu^\tau_{B_{x,\ell}}). \tag{5.2}$$

Hence to bound $c_{\mathrm{sob}}(\mu_T^\eta)$ it suffices to show that, for some constant ℓ, $\mathrm{Ent}_T^\eta(f) \leq \mathrm{const} \times \mathcal{E}_\ell(f)$ for all $f \geq 0$.

Some basic properties of variance and entropy

We record here some basic properties of variance and entropy that are repeatedly used in the proofs below:

(i) For $B \subseteq A \subseteq T$,

$$\mathrm{Var}_A^\tau(f) = \mu_A^\tau[\mathrm{Var}_B(f)] + \mathrm{Var}_A^\tau[\mu_B(f)]. \tag{5.3}$$

This equation expresses a decomposition of variance into the local conditional variance in B and the variance of the projection outside B.

(ii) If $A = \bigcup_i A_i$ for disjoint A_i, and the Gibbs distribution μ_A^τ is the product of its marginals over the A_i, then for any function f,

$$\mathrm{Var}_A^\tau(f) \leq \sum_i \mu_A^\tau[\mathrm{Var}_{A_i}(f)]. \tag{5.4}$$

(iii) For any two subsets $A, B \subseteq T$ such that $(\partial A) \cap B = \emptyset$, and for any function f,

$$\mu_T^\eta[\mathrm{Var}_A(\mu_B(f))] \leq \mu_T^\eta[\mathrm{Var}_A(\mu_{A\cap B}(f))]. \tag{5.5}$$

All three properties (i), (ii) and (iii) also hold with Var replaced by Ent. The proofs of properties (i) and (iii) are given (for completeness) in Section A.1 in the Appendix. Property (ii) is well known and also follows from our discussion of the decomposition of variance and entropy given in Sections A.2 and A.3 respectively.

Mixing in space

We are now ready to state our spatial mixing conditions, first for the variance and then for the entropy. Recall that T_x is the maximal subtree rooted at x, and that $\widetilde{T_x} \equiv T_x \setminus \{x\}$ is the subtree T_x excluding its root.

Definition 5.7 [Variance Mixing] *We say that μ_T^η satisfies $\mathrm{VM}(\ell,\varepsilon)$ if for every $x \in T$, any $\tau \in \Omega_T^\eta$ and any function f that does not depend on $B_{x,\ell}$,*

$$\mathrm{Var}^\tau_{T_x}[\mu_{\widetilde{T_x}}(f)] \;\le\; \varepsilon \cdot \mathrm{Var}^\tau_{T_x}(f).$$

Let us briefly discuss the above condition. Essentially, $\varepsilon = \varepsilon(\ell)$ gives the rate of decay with distance ℓ of point-to-set correlations. To see this, note that the l.h.s. $\mathrm{Var}^\tau_{T_x}[\mu_{\widetilde{T_x}}(f)]$ is the variance of the *projection* of f onto the root x of T_x, which is at distance ℓ from the sites on which f depends. It is also worth noting that the required uniformity in τ in VM is not very restrictive: since the distribution $\mu^\tau_{T_x}$ depends only on the restriction of τ to the boundary of T_x, and since $\tau \in \Omega_T^\eta$ (i.e., τ agrees with η on ∂T and therefore on the bottom boundary of T_x), the only freedom left in choosing τ is in choosing the spin of the parent of x. Thus, VM is essentially a property of the distribution induced by the boundary condition η. It is this lack of uniformity (i.e., the fact that we need not verify VM for other boundary conditions) that makes it flexible enough for our applications. Notice also that this means that VM with ε exponentially small in ℓ roughly corresponds to exponential decay of correlations as in Definition 2.5. (The correspondence is not exact because VM — when translated into the terminology of Definition 2.5 — only considers the case in which one of the functions depends on the spin at a single site.)

As the following theorem states, if $\mathrm{VM}(\ell,\varepsilon)$ holds with $\varepsilon \approx \frac{1}{2\ell}$, then we get a lower bound on c_{gap}.

Theorem 5.8 *For any ℓ and $\delta > 0$, if μ_T^η satisfies $\mathrm{VM}(\ell,(1-\delta)/2(\ell+1-\delta))$ then $\mathrm{Var}^\eta_T(f) \le \frac{3}{\delta}\cdot\mathcal{D}_\ell(f)$ for all f. In particular, if VM with the above parameters holds for some fixed ℓ and $\delta > 0$, for all μ_T^η with η fixed and T an arbitrary full subtree, then c_{gap} is bounded for η. Conversely, if c_{gap} is bounded for η then for some constants $c,\vartheta > 0$ and all T, μ_T^η satisfies $\mathrm{VM}(\ell, ce^{-\vartheta\ell})$ for all ℓ.*

Notice that in $\mathrm{VM}(\ell,\epsilon(\ell))$, $\epsilon(\ell)$ expresses the rate of decay of variance in space and thus Theorem 5.8 establishes a correspondence between the rate of decay of variance in space and that in time (i.e., c_{gap}). In particular, this yields an equivalence between certain notions (based on variance) of spatial and temporal mixing, as discussed in a more general setting in Chapter 2. Note crucially that the equivalence is specific to the boundary condition η.

Remark: The (easier) second part of the theorem was already proved in [KMP01], where it was shown that, for general nearest-neighbor spin systems on any bounded degree graph, if $c_{\mathrm{gap}}(\mu_T^\eta)$

is bounded independently of the volume n of T then μ_T^η exhibits exponential decay of correlations (i.e., $\mathrm{VM}(\ell, e^{-\Theta(\ell)})$ holds for all ℓ). The authors of [KMP01] posed the question of whether the converse is also true. Theorem 5.8 answers this question affirmatively when the graph is a tree. In fact, as is apparent from the above theorem, the decay of correlations on a tree is either slower than linear or exponentially fast.

The analogous mixing condition for entropy and the log-Sobolev constant is the following:

Definition 5.9 [Entropy Mixing] *We say that μ_T^η satisfies $\mathrm{EM}(\ell, \varepsilon)$ if for every $x \in T$, any $\tau \in \Omega_T^\eta$ and any non-negative function f that does not depend on $B_{x,\ell}$,*

$$\mathrm{Ent}_{T_x}^\tau[\mu_{\widetilde{T_x}}(f)] \;\leq\; \varepsilon \cdot \mathrm{Ent}_{T_x}^\tau(f).$$

Before stating the analog of Theorem 5.8 relating c_{sob} to EM, we need to define one more constant. Let $p_{\min} = \min_{x,\tau,s} \left\{ p = \mu_{T_x}^\tau(\sigma_x = s) : p > 0 \right\}$, i.e., $p_{\min}$ is the minimum non-zero probability of any spin value at any site $x \in T$ with any boundary condition at the parent of x. It is easy to see that $p_{\min}$ is bounded below by a constant that depends only on b and the potentials of system.

Theorem 5.10 *For any ℓ and $\delta > 0$, if μ_T^η satisfies $\mathrm{EM}(\ell, [(1-\delta)p_{\min}/(\ell+1-\delta)]^2)$ then $\mathrm{Ent}_T^\eta(f) \leq \frac{2}{\delta} \cdot \mathcal{E}_\ell(f)$ for all $f \geq 0$. In particular, if EM with the above parameters holds for some fixed ℓ and $\delta > 0$, for all μ_T^η with η fixed and T an arbitrary full subtree, then c_{sob} is bounded for η. Conversely, if c_{sob} is bounded for η then for some constants $c, \vartheta > 0$ and all T, μ_T^η satisfies $\mathrm{EM}(\ell, ce^{-\vartheta\ell})$ for all ℓ.*

Again, this theorem establishes a correspondence between the rates of decay of entropy in space and in time. We also notice that weak spatial mixing as in Definition 2.3 implies $\mathrm{EM}(\ell, ce^{-\Theta(\ell)})$ (and also $\mathrm{VM}(\ell, ce^{-\Theta(\ell)})$) for all boundary conditions. (This easily follows from the discussion in Appendix A.) Thus, on a tree, weak spatial mixing implies bounded c_{sob} (and thus optimal temporal mixing) uniformly in the boundary condition. The latter implication is similar to the one discussed in Chapter 4 for systems on the integer lattice, where optimal temporal mixing uniformly in the boundary condition follows from *strong* spatial mixing. (In contrast to trees, on the integer lattice the latter condition is necessary for optimal temporal mixing.)

We turn now to the proofs of Theorems 5.8 and 5.10. For this purpose, it is convenient to work with spatial mixing conditions that are somewhat more involved than VM

and EM. The main difference is that we want to allow for functions that may depend on $B_{x,\ell}$ (the first ℓ levels of T_x) and thus need to introduce a term for this dependency. The modified conditions express the property that the variance (entropy) of the projection of any function f onto the root x of T_x can be bounded up to a constant factor by the local variance (entropy) of f in $B_{x,\ell}$, plus a negligible factor times the local variance (entropy) of f in $\widetilde{T_x}$. In Appendix A we show that the modified conditions (with appropriate parameters) can be deduced from VM and EM respectively. Specializing those results to the setting here gives:

Lemma 5.11 *The following two implications hold:*

(i) *For any $\varepsilon < \frac{1}{2}$, if μ_T^η satisfies $\mathrm{VM}(\ell, \varepsilon)$ then for every $x \in T$, any $\tau \in \Omega_T^\eta$ and any function f we have $\mathrm{Var}^\tau_{T_x}[\mu_{\widetilde{T_x}}(f)] \leq \frac{2-\varepsilon'}{1-\varepsilon'} \cdot \mu^\tau_{T_x}[\mathrm{Var}_{B_{x,\ell}}(f)] + \frac{\varepsilon'}{1-\varepsilon'} \cdot \mu^\tau_{T_x}[\mathrm{Var}_{\widetilde{T_x}}(f)]$, with $\varepsilon' = 2\varepsilon$.*

(ii) *For any $\varepsilon < p_{\min}^2$, if μ_T^η satisfies $\mathrm{EM}(\ell, \varepsilon)$ then for every $x \in T$, any $\tau \in \Omega_T^\eta$ and any function $f \geq 0$ we have $\mathrm{Ent}^\tau_{T_x}[\mu_{\widetilde{T_x}}(f)] \leq \frac{1}{1-\varepsilon'} \cdot \mu^\tau_{T_x}[\mathrm{Ent}_{B_{x,\ell}}(f)] + \frac{\varepsilon'}{1-\varepsilon'} \cdot \mu^\tau_{T_x}[\mathrm{Ent}_{\widetilde{T_x}}(f)]$, with $\varepsilon' = \frac{\sqrt{\varepsilon}}{p_{\min}}$.*

The two implications are special cases of Theorems A.5 and A.8 respectively. For implication (ii), notice that Theorem A.8 gives a bound on $\mathrm{Ent}^\tau_{T_x}(f)$, and we replaced the latter with $\mathrm{Ent}^\tau_{T_x}[\mu_{\widetilde{T_x}}(f)] + \mu^\tau_{T_x}[\mathrm{Ent}_{\widetilde{T_x}}(f)]$, since this sum equals $\mathrm{Ent}^\tau_{T_x}(f)$ by the entropy version of (5.3). We refer to Appendix A for the general theorems and their proofs.

Remark: We note that, with some extra work, part (ii) of Lemma 5.11 (i.e., Theorem A.8) can be improved to hold with $\varepsilon' = c\varepsilon$, where c is a constant that depends only on $p_{\min}$. We content ourselves with the weaker bound because it is simpler to prove while still enough for our applications.

We can now prove Theorems 5.8 and 5.10 by working with the modified spatial mixing conditions of Lemma 5.11.

Proof of Theorems 5.8 and 5.10: Here we only prove the forward direction of both theorems. The reverse direction of Theorem 5.8 was proved in [KMP01] (for general graphs), as already mentioned above. The reverse direction of Theorem 5.10 is deferred to Appendix B, where it is also proved for general graphs. The reason for this is that the proof uses machinery developed in Appendix A, and that the reverse direction is not needed for the rest of the development in this chapter.

The main step in the proof of the forward direction is to show the following claim:

Claim 5.12 *If for every $x \in T$, any $\tau \in \Omega_T^\eta$ and any function f we have*

$$\mathrm{Var}_{T_x}^\tau[\mu_{\widetilde{T_x}}(f)] \;\le\; c \cdot \mu_{T_x}^\tau[\mathrm{Var}_{B_{x,\ell}}(f)] \;+\; \left(\frac{1-\delta}{\ell}\right) \cdot \mu_{T_x}^\tau[\mathrm{Var}_{\widetilde{T_x}}(f)],$$

then $\mathrm{Var}(f) \le \frac{c}{\delta} \cdot \mathcal{D}_\ell(f)$ for all f. The same implication holds when Var is replaced by Ent, $\mathcal{D}_\ell$ is replaced by $\mathcal{E}_\ell$ and the function f is restricted to be non-negative.

Observe that the hypothesis of Theorem 5.8 together with part (i) of Lemma 5.11 establishes the hypothesis of Claim 5.12 with $c \le 3$, and similarly, the hypothesis of Theorem 5.10 together with part (ii) of Lemma 5.11 establishes the hypothesis of Claim 5.12 (after the necessary replacement of symbols) with $c \le 2$.

It therefore suffices to prove Claim 5.12. We prove only the formulation with Var and $\mathcal{D}_\ell$ since the proof for the formulation with Ent and $\mathcal{E}_\ell$ is identical once we make the same replacements in the text of the proof. As will be clear below, the proof uses only properties (5.3), (5.4) and (5.5), which are common to both Var and Ent.

Consider an arbitrary function $f : \Omega \to \mathbb{R}$. Our first goal is to relate $\mathrm{Var}_T^\eta(f)$ to the projections $\mathrm{Var}_{T_x}^\tau[\mu_{\widetilde{T_x}}(f)]$ for $x \in T$, so that we can apply the spatial mixing condition of the hypothesis. Recall that T has $k+1$ levels, and define the increasing sequence $\emptyset = F_0 \subset F_1 \subset \ldots \subset F_{k+1} = T$, where F_i consists of all sites in the lowest i levels of T. Thus F_i is a forest of height $i-1$. Using (5.3) recursively, and the facts that $\mu_{F_{i+1}}(\mu_{F_i}(f)) = \mu_{F_{i+1}}(f)$ and $\mu_{F_0}(f) = f$, we obtain

$$\begin{aligned}
\mathrm{Var}_T^\eta(f) &= \mu_T^\eta[\mathrm{Var}_{F_1}(f)] + \mathrm{Var}_T^\eta[\mu_{F_1}(f)] \\
&= \mu_T^\eta[\mathrm{Var}_{F_1}(f)] + \mu_T^\eta[\mathrm{Var}_{F_2}(\mu_{F_1}(f))] + \mathrm{Var}_T^\eta[\mu_{F_2}(\mu_{F_1}(f))] \\
&\;\;\vdots \\
&= \sum_{i=1}^{k+1} \mu_T^\eta[\mathrm{Var}_{F_i}(\mu_{F_{i-1}}(f))].
\end{aligned}$$

Now a fundamental property of nearest-neighbor interaction models on a tree is that, given the configuration on $T \setminus F_i$, the Gibbs distribution on F_i becomes a product of the marginals on the subtrees rooted at the sites $x \in F_i \setminus F_{i-1}$. Using inequality (5.4) for the variance of a product measure, we therefore have that

$$\mathrm{Var}_T^\eta(f) \le \sum_{i=1}^{k+1} \sum_{x \in F_i \setminus F_{i-1}} \mu_T^\eta[\mathrm{Var}_{T_x}(\mu_{F_{i-1}}(f))] \le \sum_{x \in T} \mu_T^\eta[\mathrm{Var}_{T_x}(\mu_{\widetilde{T_x}}(f))], \qquad (5.6)$$

where in the second inequality we used the convexity of the variance as in (5.5).

Notice that so far we have not used the spatial mixing condition in the hypothesis of Claim 5.12, but only a natural martingale structure induced by the tree. Let us denote the final sum in (5.6) by $\mathrm{Pvar}_T^\eta(f)$. In order to bound c_{gap}, we need to compare the projection terms $\mathrm{Var}_{T_x}(\mu_{\widetilde{T_x}}(f))$ in $\mathrm{Pvar}_T^\eta(f)$ with the local conditional variance terms in $\mathcal{D}_\ell(f)$. For example, notice that if μ_T^η were the product of its single-site marginals then $\mathrm{Var}_{T_x}(\mu_{\widetilde{T_x}}(f)) \leq \mu_{T_x}[\mathrm{Var}_x(f)]$ and $c_{\mathrm{gap}} = 1$. However, in general the variance of the projection on x may also involve terms which depend on other sites, and may lead to a factor that grows with the size of T_x. We will use the spatial mixing condition in order to preclude the latter possibility. Specifically, we show that if for every $x \in T$, any $\tau \in \Omega_T^\eta$ and any function g, $\mathrm{Var}_{T_x}^\tau[\mu_{\widetilde{T_x}}(g)] \leq c \cdot \mu_{T_x}^\tau[\mathrm{Var}_{B_{x,\ell}}(g)] + \varepsilon \cdot \mu_{T_x}^\tau[\mathrm{Var}_{\widetilde{T_x}}(g)]$ then for every $x \in T$ and $\tau \in \Omega$,

$$\mathrm{Var}_{T_x}^\tau[\mu_{\widetilde{T_x}}(f)] \leq c \cdot \mu_{T_x}^\tau[\mathrm{Var}_{B_x}(f)] + \varepsilon \cdot \sum_{y \in B_x \cup \widetilde{\partial} B_x, y \neq x} \mu_{T_x}^\tau[\mathrm{Var}_{T_y}(\mu_{\widetilde{T_y}}(f))], \tag{5.7}$$

where we have abbreviated $B_{x,\ell}$ to B_x and $\widetilde{\partial} B_x$ stands for the boundary of B_x excluding the parent of x, i.e., the bottom boundary of B_x. Notice that the last term in (5.7) is relevant only when x is at distance at least ℓ from the bottom of T. When x belongs to one of the ℓ lowest levels of T then $T_x = B_x$, and thus trivially $\mathrm{Var}_{T_x}^\tau[\mu_{\widetilde{T_x}}(f)] \leq \mu_{T_x}^\tau[\mathrm{Var}_{B_x}(f)]$.

Let us assume (5.7) for now and conclude the proof of the theorem. Applying (5.7) for every x and τ, and using the hypothesis that $\varepsilon = \frac{1-\delta}{\ell}$ and the fact that each site appears in at most ℓ blocks, we get

$$\begin{aligned}
\mathrm{Pvar}_T^\eta(f) &\leq c \cdot \mathcal{D}_\ell(f) + \varepsilon \cdot \sum_{x \in T} \sum_{y \in B_x \cup \widetilde{\partial} B_x, y \neq x} \mu_T^\eta[\mathrm{Var}_{T_y}(\mu_{\widetilde{T_y}}(f))] \\
&\leq c \cdot \mathcal{D}_\ell(f) + \varepsilon \ell \cdot \sum_{y \in T} \mu_T^\eta[\mathrm{Var}_{T_y}(\mu_{\widetilde{T_y}}(f))] \\
&= c \cdot \mathcal{D}_\ell(f) + (1-\delta)\mathrm{Pvar}_T^\eta(f),
\end{aligned}$$

and hence

$$\mathrm{Var}_T^\eta(f) \leq \mathrm{Pvar}_T^\eta(f) \leq \frac{c}{\delta} \cdot \mathcal{D}_\ell(f),$$

proving Claim 5.12. We now return to proving (5.7).

Let $g = \mu_{T_x \setminus (B_x \cup \widetilde{\partial} B_x)}(f)$. Once we notice that $\mu_{\widetilde{T_x}}(f) = \mu_{\widetilde{T_x}}(g)$, we can use the spatial mixing assumption that precedes (5.7) to deduce

$$\begin{aligned}
\mathrm{Var}_{T_x}^\tau[\mu_{\widetilde{T_x}}(f)] &\leq c \cdot \mu_{T_x}^\tau[\mathrm{Var}_{B_x}(g)] + \varepsilon \cdot \mu_{T_x}^\tau[\mathrm{Var}_{\widetilde{T_x}}(g)] \\
&\leq c \cdot \mu_{T_x}^\tau[\mathrm{Var}_{B_x}(f)] + \varepsilon \cdot \mu_{T_x}^\tau[\mathrm{Var}_{\widetilde{T_x}}(g)],
\end{aligned}$$

where we used (5.5) for the second inequality. We will be done once we show that

$$\mu^{\tau}_{T_x}[\mathrm{Var}_{\widetilde{T_x}}(g)] \;\leq \sum_{y\in B_x\cup\widetilde{\partial}B_x, y\neq x} \mu^{\tau}_{T_x}[\mathrm{Var}_{T_y}(\mu_{\widetilde{T_y}}(f))]. \tag{5.8}$$

But (5.8) follows from a similar argument to that used earlier to show $\mathrm{Var}^{\eta}_{T}(f) \leq \mathrm{Pvar}^{\eta}_{T}(f)$, starting from the fact that $g = \mu_{F'_d}(f)$, where the forests F'_i are defined analogously to the F_i earlier but restricted to the subtree T_x, and $d = \mathrm{height}(x) - \ell$. We omit the details.

This concludes the proof of Claim 5.12, and thus of Theorems 5.8 and 5.10. □

5.2.2 Establishing VM through disagreement percolation

We now move on to the second part of our framework, where we show that the conditions on κ and γ in parts (i) and (ii) of Theorem 5.6 imply the Variance Mixing condition VM and the Entropy Mixing condition EM respectively, with ε exponentially small in ℓ. The analysis is carried out separately for each of the two cases, and we first give the analysis for part (i) and VM. The main result in this section reads as follows.

Theorem 5.13 *Any Gibbs distribution μ^{η}_{T} satisfies $\mathrm{VM}(\ell, p_{\min}^{-1}(\gamma\kappa b)^{\ell})$ for all ℓ, where κ and γ are the constants associated with the sequence $\{\mu^{\eta}_{T}\}$ as specified in Definition 5.4. In particular, if $\gamma\kappa b < 1$ then there exists a constant $\vartheta > 0$ such that, for every T, μ^{η}_{T} satisfies $\mathrm{VM}(\ell, p_{\min}^{-1}e^{-\vartheta\ell})$ for all ℓ.*

Notice that part (i) of Theorem 5.6 follows by combining Theorems 5.13 and 5.8. Specifically, Theorem 5.8 tells us that in order to obtain bounded c_{gap}, it is enough to establish the Variance Mixing condition $\mathrm{VM}(\ell, \varepsilon)$ with $\varepsilon = (1-\delta)/2(\ell+1-\delta)$, for some constants $\ell, \delta > 0$ independent of the size of T. However, using Theorem 5.13, the hypothesis in part (i) of Theorem 5.6 implies that VM with the above parameters indeed holds for large enough ℓ and some $\delta > 0$. We conclude that the same hypothesis implies bounded c_{gap}, as required.

Remarks:

- The factor $p_{\min}^{-1}$ in Theorem 5.13 can be improved to $q-1$, where $q = |\mathcal{S}|$ is the size of the spin space $\mathcal{S}$, but the constant $p_{\min}^{-1}$ is good enough for our purposes here, and the proof for this factor is simpler.
- The validity of VM, i.e, the decay of point-to-set correlations, is of interest independently of its implication for the spectral gap (an implication which is new here): e.g., it is closely

related to the purity of the infinite volume Gibbs measure and to bit reconstruction problems on trees [EKPS00]. Compared to other arguments for establishing VM that are based on disagreement percolation, the novelty of our argument is that it considers the product of κ and γ, rather than working with κ alone.

Proof of Theorem 5.13: The proof of the theorem is based on the disagreement percolation argument expressed in Claim 5.5. (Recall that the proof of this claim is given in Section 5.2.4.) To understand the role the claim plays in the proof here, fix T, $x \in T$ and $\tau \in \Omega_T^\eta$, and recall that $\mu^s_{\widetilde{T_x}} \equiv \mu^{\tau^{x,s}}_{\widetilde{T_x}}$ is the Gibbs distribution in which the spin at x is fixed to s. Now, let $\mu^s_\ell \equiv \mu^s_{\widetilde{T_x}}(\mu_{B_{x,\ell}})$ stand for the distribution described by the following two-stage procedure: first, choose a configuration σ from $\mu^s_{\widetilde{T_x}}$ (i.e., conditioned on s being the spin at the root x); then, choose a new configuration according to $\mu^\sigma_{B_{x,\ell}}$ (i.e., the configuration σ chosen in the first step specifies the boundary condition at the bottom of $B_{x,\ell}$ in the second step). Let $d(\ell) = \max_{s_1,s_2} \|\mu^{s_1}_\ell - \mu^{s_2}_\ell\|_x$. Thus $d(\ell)$ is an upper bound on the influence of the initial spin at x on the resulting spin at x in the above procedure. We now use Claim 5.5 to show that $d(\ell) \leq (\gamma\kappa b)^\ell$. Specifically, for all s_1 and s_1,

$$\begin{aligned}
\|\mu^{s_1}_\ell - \mu^{s_2}_\ell\|_x &= \|\mu^{s_1}_{\widetilde{T_x}}(\mu_{B_{x,\ell}}) - \mu^{s_2}_{\widetilde{T_x}}(\mu_{B_{x,\ell}})\|_x \\
&\leq \sum_{\sigma,\sigma'} \nu(\sigma,\sigma')\|\mu^\sigma_{B_{x,\ell}} - \mu^{\sigma'}_{B_{x,\ell}}\|_x \\
&\leq \sum_{\sigma,\sigma'} \nu(\sigma,\sigma')|\sigma - \sigma'|_{x,\ell} \cdot \gamma^\ell \\
&= \gamma^\ell \cdot \mathrm{E}_\nu|\sigma - \sigma'|_{x,\ell} \\
&\leq (\gamma\kappa b)^\ell,
\end{aligned}$$

where $\nu = \nu_x^{s_1,s_2}$ is the coupling of $\mu^{s_1}_{\widetilde{T_x}}$ and $\mu^{s_2}_{\widetilde{T_x}}$ from part (i) of Claim 5.5, and in the third line we used part (ii) of that claim.

The importance of $d(\ell)$ is that it is closely related to the *rate* of the Variance Mixing VM. Recall that in order to establish $\mathrm{VM}(\ell,\varepsilon)$, we need to show that for every function f that does not depend on $B_{x,\ell}$,

$$\mathrm{Var}^\tau_{T_x}[\mu_{\widetilde{T_x}}(f)] \;\leq\; \varepsilon \cdot \mathrm{Var}^\tau_{T_x}(f). \tag{5.9}$$

As we show below, (5.9) holds with $\varepsilon = d(\ell)/p_{\min}$ for every function f that does not depend on $B_{x,\ell}$. Since $d(\ell) \leq (\gamma\kappa b)^\ell$, this will complete the proof of Theorem 5.13.

Now, note that (5.9) means that projecting f onto the root (of T_x) decreases the variance by a factor ε. As is well known, in order to establish this it is enough to prove a dual contraction, i.e., to consider an arbitrary function that depends only on the spin at the root and show that, when projecting onto levels ℓ and below, the variance shrinks by a factor ε. Formally, it is enough to show that for every function g that does not depend on $\widetilde{T_x}$ [3] we have

$$\mathrm{Var}^{\tau}_{T_x}[\mu_{B_{x,\ell}}(g)] \;\le\; \varepsilon \cdot \mathrm{Var}^{\tau}_{T_x}(g). \tag{5.10}$$

To see that (5.10) implies (5.9), we refer (for completeness) to Eqn. (A.7) in Appendix A.

We now complete the proof by showing that (5.10) holds with $\varepsilon = d(\ell)/p_{\min}$. Consider an arbitrary g that does not depend on $\widetilde{T_x}$. Let $p(s) \equiv \mu^{\tau}_{T_x}(\sigma_x = s)$. We also write g^s for $g(\sigma)$, where σ is any configuration that agrees with τ outside T_x and such that $\sigma_x = s$. (This is well defined since g does not depend on $\widetilde{T_x}$). Also, let $g_{\max} = \max_s \{g^s : p(s) > 0\}$ stand for the maximum value of g, and define $g_{\min}$ similarly. Then,

$$\begin{aligned}
\mathrm{Var}^{\tau}_{T_x}[\mu_{B_{x,\ell}}(g)] &= \mathrm{Cov}^{\tau}_{T_x}[g, \mu_{B_{x,\ell}}(g)] \\
&= \mathrm{Cov}^{\tau}_{T_x}[g, \mu_{\widetilde{T_x}}(\mu_{B_{x,\ell}}(g))] \\
&= \mathrm{Cov}^{\tau}_{T_x}[g, \mu_{\ell}(g)] \\
&= \sum_{s\in\mathcal{S}} p(s)\left[g^s - \mu^{\tau}_{T_x}(g)\right]\left[\mu^s_{\ell}(g) - \mu^{\tau}_{T_x}(\mu_{\ell}(g))\right] \\
&\le \sum_{s\in\mathcal{S}} p(s)\left|g^s - \mu^{\tau}_{T_x}(g)\right| \max_{s'\in\mathcal{S}}\left|\mu^s_{\ell}(g) - \mu^{s'}_{\ell}(g)\right| \\
&\le \sum_{s\in\mathcal{S}} p(s)\left|g^s - \mu^{\tau}_{T_x}(g)\right| d(\ell)\left|g_{\max} - g_{\min}\right| \\
&\le d(\ell)\left[\left|g_{\max} - \mu^{\tau}_{T_x}(g)\right| + \left|\mu^{\tau}_{T_x}(g) - g_{\min}\right|\right] \sum_{s\in\mathcal{S}} p(s)\left|g^s - \mu^{\tau}_{T_x}(g)\right| \\
&\le \frac{d(\ell)}{p_{\min}}\left[\sum_{s\in\mathcal{S}} p(s)\left|g^s - \mu^{\tau}_{T_x}(g)\right|\right]^2 \\
&\le \frac{d(\ell)}{p_{\min}} \cdot \mathrm{Var}^{\tau}_{T_x}(g),
\end{aligned}$$

where for the second equality we used the fact that g does not depend on $\widetilde{T_x}$, and the last inequality is an application of Cauchy-Schwartz. □

[3] Effectively this means that, conditioned on the configuration outside T_x being τ, g depends only on the spin at the root x.

5.2.3 Establishing EM through disagreement percolation

In this section we complete the proof of part (ii) of Theorem 5.6, by showing that the condition on κ and γ in part (ii) implies the Entropy Mixing condition EM with ε exponentially small in ℓ. The main result in this section reads as follows.

Theorem 5.14 *Any Gibbs distribution μ_T^η satisfies $\mathrm{EM}(\ell, c(\gamma\alpha)^{\ell/5})$ for all ℓ, where $\alpha = \max\{\kappa b, 1\}$, κ and γ are the constants associated with the sequence $\{\mu_T^\eta\}$ as specified in Definition 5.4, and c is a constant that depends only on $p_{\min}$ and $(\gamma\alpha)$. In particular, if $\max\{\gamma\kappa b, \gamma\} < 1$ then there exists a constant ϑ such that, for every T, μ_T^η satisfies $\mathrm{EM}(\ell, ce^{-\vartheta\ell})$ for all ℓ.*

Notice that part (ii) of Theorem 5.6 follows by combining Theorems 5.14 and 5.10, exactly as was done for part (i) immediately following Theorem 5.13.

Proof of Theorem 5.14: Fix arbitrary T, $x \in T$ and $\tau \in \Omega_T^\eta$ for the rest of this proof. Our argument is composed of three main ingredients, and as in the proof of Theorem 5.13, here too the first ingredient is based on the bounds κ and γ give for the probability of disagreements percolating down and up the tree respectively, as expressed by Claim 5.5. However, here we will need a strengthening of part (i) of the claim. In particular, rather than just a bound on the *average* Hamming distance under the coupling given in the claim, here we will need a *strong tail bound* for this Hamming distance, i.e., we need to show that this distance is not much larger than the bound given in Claim 5.5, with very high probability. This is the content of the following lemma. Recall the notation used in Claim 5.5 and the coupling ν given in part (i) of the claim.

Lemma 5.15 *For an integer ℓ, let $\overline{\mathcal{H}}_\ell = \max \mathrm{E}_\nu |\sigma - \sigma'|_{z,i}$, where the maximum is taken over all $z \in T_x$, $s_1, s_2 \in \mathcal{S}$ and $0 \le i \le \ell$, and where $\nu \equiv \nu_z^{s_1,s_2}$ is the coupling of $\mu_{\widetilde{T}_z}^{s_1}$ and $\mu_{\widetilde{T}_z}^{s_2}$. Then for all s_1, s_2 and every $C > 0$,*

$$\Pr_\nu \left[|\sigma - \sigma'|_{x,\ell} > C\overline{\mathcal{H}}_\ell \right] \;\le\; e^{\frac{1}{\ell+1}\left(1 - \frac{C}{2e}\right)},$$

where $\nu \equiv \nu_x^{s_1,s_2}$.

Notice that Claim 5.5 asserted the existence of a coupling ν with certain properties without specifying ν explicitly. However, a specific coupling is constructed in the proof of this claim, and it is this coupling that Lemma 5.15 refers to. The proof of the lemma goes by standard

arguments from the analysis of branching processes, and is given in Section 5.2.4 following the proof of Claim 5.5.

In order to see the role the above lemma plays in the proof of Theorem 5.14, we need to introduce new notation. Recall that $p(s) \equiv \mu^{\tau}_{T_x}(\sigma_x = s)$, and write $\mathcal{S}^+ \equiv \{s : p(s) > 0\}$ for the support of p. For $s \in \mathcal{S}^+$ and $\sigma \in \Omega^{\tau}_{T_x}$, define

$$g_s(\sigma) = \frac{\mu^{s}_{\widetilde{T}_x}(\sigma)}{\mu^{\tau}_{T_x}(\sigma)} = \begin{cases} 1/p(s) & \text{if } \sigma_x = s, \\ 0 & \text{otherwise.} \end{cases}$$

The key quantity we will work with in the sequel is the following:

$$g^{(\ell)}_s = \mu_{B_{x,\ell}}(g_s).$$

Note that $g^{(\ell)}_s(\sigma)$ depends only on the spins in $\widetilde{\partial} B_{x,\ell}$. Indeed, let $\sigma_{x,\ell}$ stand for the restriction of σ to $\widetilde{\partial} B_{x,\ell}$, i.e., to the sites at distance ℓ below x. It is easy to verify that $g^{(\ell)}_s(\sigma)$ is equal to $\frac{\mu^{s}_{\widetilde{T}_x}(\sigma_{x,\ell})}{\mu^{\tau}_{T_x}(\sigma_{x,\ell})}$. Thus, for a given configuration σ, $g^{(\ell)}_s(\sigma)$ is the ratio of the probabilities of seeing the spins of σ at level ℓ below the root x when the spin at x is s and when there is no condition on the spin at x, respectively. Now, recall that in the proof of Theorem 5.13, the first ingredient for establishing VM was combining parts (i) and (ii) of Claim 5.5 to bound the probability of disagreement $d(\ell)$ in the two-stage procedure described there. Here we take a similar step by combining the two parts to get the following corollary of the strong tail bound in Lemma 5.15.

Corollary 5.16 *For every $C > 0$, $s \in \mathcal{S}^+$, and all (s_1, s_2),*

$$\Pr_{\nu}\left[\left|g^{(\ell)}_s(\sigma) - g^{(\ell)}_s(\sigma')\right| > C(\gamma\alpha)^{\ell}\right] \;\le\; e^{\frac{1}{\ell+1}\left(1 - \frac{p_{\min} C}{2e}\right)},$$

where $\nu \equiv \nu_x^{s_1,s_2}$ is the coupling given in Claim 5.5.

Proof: It is enough to show that

$$|g^{(\ell)}_s(\sigma) - g^{(\ell)}_s(\sigma')| \le \frac{\gamma^{\ell}}{p_{\min}} \cdot |\sigma - \sigma'|_{x,\ell} \tag{5.11}$$

since we can then apply Lemma 5.15 with C replaced by $p_{\min} C$, and by observing that $\overline{\mathcal{H}}_{\ell} \le \alpha^{\ell}$ by part (i) of Claim 5.5. On the other hand, (5.11) follows from part (ii) of Claim 5.5 once we recall that $g^{(\ell)}_s(\sigma) = \mu^{\sigma}_{B_{x,\ell}}(g_s)$ and that g_s depends only on the spin at the root x, implying that $|g^{(\ell)}_s(\sigma) - g^{(\ell)}_s(\sigma')| \le \|\mu^{\sigma}_{B_{x,\ell}} - \mu^{\sigma'}_{B_{x,\ell}}\|_x \cdot \|g_s\|_{\infty} \le \gamma^{\ell} |\sigma - \sigma'|_{x,\ell} / p_{\min}$.

□

Continuing with the proof of Theorem 5.14, the second ingredient is translating the above tail bound for the difference between the values of $g_s^{(\ell)}$ under the coupling ν, to a strong concentration of $g_s^{(\ell)}$ around its mean value 1 under the Gibbs distribution $\mu_{T_x}^{\tau}$. This is the content of the following lemma.

Lemma 5.17 *Let s be a spin value. If there exists $\delta > 0$ and a coupling ν of $\mu_{\widetilde{T_x}}^{s}$ and $\mu_{T_x}^{\tau}$ such that*

$$\Pr_{\nu}\left[\left|g_s^{(\ell)}(\sigma) - g_s^{(\ell)}(\sigma')\right| > \delta^3\right] \;\leq\; \delta e^{-1/\delta},$$

then

$$\mu_{T_x}^{\tau}[|g_s^{(\ell)} - 1| > 2\delta] \;\leq\; 4e^{-1/\delta}.$$

Let us defer the proof of Lemma 5.17 for now, and move on to complete the proof of Theorem 5.14.

Notice that strong concentration of $g_s^{(\ell)}$ is a strong form of lack of correlation between the spin at the root and the configuration ℓ levels below, i.e., it means that the spin σ_x at the root is very marginally affected by conditioning on the configuration $\sigma_{x,\ell}$, for almost all $\sigma_{x,\ell}$. The third and final ingredient we need in order to prove Theorem 5.14 is a translation of this strong lack of correlation to EM, which is an alternative form of lack of correlation between σ_x and $\sigma_{x,\ell}$. Such a translation is given in a more general setting in Theorem A.10, and we cite below a specialization of it to the scenario here.

Lemma 5.18 *There exists a numerical constant c such that, for any $\delta \geq 0$, if*

$$\mu_{T_x}^{\tau}\left[|g_s^{(\ell)} - 1| > \delta\right] \;\leq\; e^{-2/\delta} \tag{5.12}$$

for all $s \in \mathcal{S}^+$, then we have $\mathrm{Ent}_{T_x}^{\tau}\left[\mu_{\widetilde{T_x}}(f)\right] \leq c p_{\min}^{-2}\delta \mathrm{Ent}_{T_x}^{\tau}(f)$ for any non-negative function f that does not depend on $B_{x,\ell}$; in particular, $\mathrm{EM}(\ell, cp_{\min}^{-2}\delta)$ holds.

We refer to Theorem A.10 in Appendix A.3 for a proof.

Combining Corollary 5.16, Lemma 5.17 and Lemma 5.18 proves Theorem 5.14 as we now explain. First, notice that w.l.o.g. we can assume that $\gamma\alpha < 1$, because otherwise the theorem is trivial since $\mathrm{EM}(\ell, 1)$ always holds. We can also assume that $\gamma > 0$, because $\gamma = 0$ means that the spins are completely independent of each other, and hence $\mathrm{EM}(\ell, 0)$ holds. Now observe that, since $\mu_{T_x}^{\tau}$ is a convex combination of the distributions $\mu_{\widetilde{T_x}}^{s'}$ as s' varies, there exists a coupling ν of $\mu_{\widetilde{T_x}}^{s}$ and $\mu_{T_x}^{\tau}$ that satisfies the bounds given in Corollary 5.16. Thus, by applying the latter with $C = (\gamma\alpha)^{0.28\ell}$, we get that the hypothesis of

Lemma 5.17 holds with $\delta = (\gamma\alpha)^{0.24\ell}$ for all large enough ℓ, and therefore that the hypothesis of Lemma 5.18 holds with, for example, $\delta = (\gamma\alpha)^{0.2\ell}$ for all large enough ℓ. Finally, by applying Lemma 5.18 with this value of δ, we get that $\mathrm{EM}(\ell, c(\gamma\alpha)^{0.2\ell})$ holds for some constant c and all ℓ, thus completing the proof of Theorem 5.14. □

Finally, we go back and supply the missing proof of Lemma 5.17.

Proof of Lemma 5.17: We will show only that under the hypothesis of the lemma

$$\mu^{\tau}_{T_x}\left[g_s^{(\ell)} - 1 > 2\delta\right] \leq 2e^{-1/\delta}, \tag{5.13}$$

since the same bound on the negative tail can be achieved by an analogous argument.

By the hypothesis, for every $\varepsilon > 0$,

$$\mu^{s}_{\widetilde{T_x}}\left[g_s^{(\ell)} - 1 > \varepsilon\right] \;\leq\; \mu^{\tau}_{T_x}\left[g_s^{(\ell)} - 1 > \varepsilon - \delta^3\right] + \delta e^{-1/\delta}. \tag{5.14}$$

Next, we notice that by definition of $g_s^{(\ell)}$,

$$\mu^{s}_{\widetilde{T_x}}\left[g_s^{(\ell)} - 1 > \varepsilon\right] \;\geq\; (1+\varepsilon)\mu^{\tau}_{T_x}\left[g_s^{(\ell)} - 1 > \varepsilon\right]. \tag{5.15}$$

Combining (5.14) and (5.15) we get that, for every $\varepsilon > 0$,

$$\mu^{\tau}_{T_x}\left[g_s^{(\ell)} - 1 > \varepsilon\right] \;\leq\; \left(\frac{1}{1+\varepsilon}\right)\left(\mu^{\tau}_{T_x}\left[g_s^{(\ell)} - 1 > \varepsilon - \delta^3\right] + \delta e^{-1/\delta}\right). \tag{5.16}$$

This immediately yields that, for every non-negative integer k and $\varepsilon > 0$,

$$\mu^{\tau}_{T_x}\left[g_s^{(\ell)} - 1 > \varepsilon + k\delta^3\right] \;\leq\; (1+\varepsilon)^{-(k+1)} + \left(\frac{1}{\varepsilon}\right)\delta e^{-1/\delta}, \tag{5.17}$$

where we applied (5.16) $k+1$ times, each time increasing ε by δ^3.

Inequality (5.13) then follows by applying (5.17) with $\varepsilon = \delta$ and $k = \lfloor (1/\delta)^2 \rfloor$.

□

5.2.4 Bounds on disagreement percolation

In this section, we give the proofs of Claim 5.5 and Lemma 5.15, which establish bounds on disagreement percolation in terms of κ and γ.

Proof of Claim 5.5: The proof makes use of a standard recursive coupling along paths in the tree (as in, e.g., [KMP01]). We start with part (i), i.e., constructing a coupling $\nu_x^{s_1,s_2}$ of $\mu_{\widetilde{T}_x}^{s_1}$ and $\mu_{\widetilde{T}_x}^{s_2}$ such that $\mathrm{E}_\nu|\sigma - \sigma'|_{x,l} \leq (\kappa b)^\ell$. Since the underlying graph is a tree, we can couple $\mu_{\widetilde{T}_x}^{s_1}$ and $\mu_{\widetilde{T}_x}^{s_2}$ recursively. This goes as follows. First, given the spin at x the distributions on T_z (where z ranges over the children of x) are all independent of each other, so we can couple the projections on the T_z's independently. Then, we couple the two projections on T_z by first coupling the spin at z using the optimal coupling (the one that achieves the variation distance) of the marginal distributions on the spin at z. Thus, the spins at z disagree with probability at most κ. Once a coupled pair of spins at z is chosen, we continue as follows: if the spins at z agree then we can make the configurations in $\widetilde{T_z}$ equal with probability 1 (because the two boundary conditions are the same); if the spins at z differ, say the first is s_1' and the second is s_2', then we recursively couple $\mu_{\widetilde{T}_z}^{s_1'}$ and $\mu_{\widetilde{T}_z}^{s_2'}$ (i.e., using the coupling $\nu_z^{s_1',s_2'}$). We let $\nu \equiv \nu_x^{s_1,s_2}$ be the resulting coupling of $\mu_{\widetilde{T}_x}^{s_1}$ and $\mu_{\widetilde{T}_x}^{s_1}$, and notice that $\mathrm{E}_\nu|\sigma - \sigma'|_{x,l} \leq (\kappa b)^\ell$ since for every site y at distance ℓ below x the probability that the two coupled spins at y disagree is at most κ^ℓ.

We go on to prove part (ii) of Claim 5.5. First, by writing a telescopic sum and applying the triangle inequality we get that

$$\|\mu_{B_{x,\ell}}^{\tau} - \mu_{B_{x,\ell}}^{\tau'}\|_x \;\leq\; \sum_{i=1}^{d} \|\mu_{B_{x,\ell}}^{\tau^{(i-1)}} - \mu_{B_{x,\ell}}^{\tau^{(i)}}\|_x \;\;,$$

where $d = |\tau - \tau'|_{x,\ell}$ and the sequence of configurations $\tau^{(i)}$ is a site-by-site interpolation of the differences between τ and τ' in $\widetilde{\partial} B_{x,\ell}$. (It suffices to interpolate only over the differences in $\widetilde{\partial} B_{x,\ell}$ since the distribution $\mu_{B_{x,\ell}}^{\tau}$ depends only on the configuration in $\partial B_{x,\ell}$ and since τ and τ' agree on the parent of x.)

It is now enough to show that $\|\mu_{B_{x,\ell}}^{\tau^{w,s_1}} - \mu_{B_{x,\ell}}^{\tau^{w,s_2}}\|_x \leq \gamma^\ell$ for all τ, $w \in \widetilde{\partial} B_{x,\ell}$, and $s_1, s_2 \in \mathcal{S}$, where we recall that $\tau^{w,s}$ denotes the configuration τ with the spin at w set to s. This, however, follows by a coupling argument as before, where this time we couple recursively along the path from w to x (i.e., up the tree). Specifically, suppose by induction that in our coupling there is already a path of disagreement going from w to y, where y is some site on the path from w to x. Let z denote the parent of y. At the next step we choose a coupled pair of spins at z from the two distributions $\mu_A^{\tau_1}$ and $\mu_A^{\tau_2}$, where the subset A is $B_{x,\ell}$ excluding the path from w to y (see Figure 5.2), and the configurations τ_1, τ_2 are equal to τ except that the configuration along the path from w to y is that chosen by the

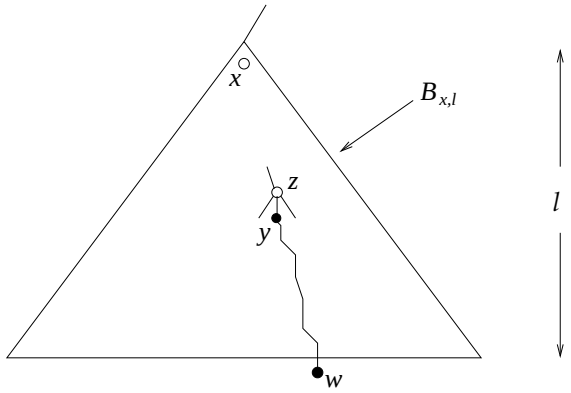

Figure 5.2: Coupling up the tree: given a path of disagreement from w to y, a coupled spin at z is chosen from a coupling of the conditional distributions on A, where A is $B_{x,\ell}$ with the path from w to y excluded.

previous steps of the recursion, in each copy respectively. Notice, however, that once the spin at y is fixed, the rest of the configuration on the path from w to y has no influence on the distribution of the spin at z. Therefore, w.l.o.g. we can assume that the coupled pair of spins at z is chosen from the distributions $\mu_A^{\tau^{y,s_1'}}$ and $\mu_A^{\tau^{y,s_2'}}$ (using an optimal coupling for the projections onto the spin at z), where (s_1', s_2') is the pair of spins at y chosen in the previous step of the recursion. Thus, the probability of disagreement at z given the disagreement at y is $\|\mu_A^{\tau^{y,s_1'}} - \mu_A^{\tau^{y,s_2'}}\|_z \leq \gamma$, by definition of γ. Now, if the resulting spins at z agree then the spins on the rest of the path to x are coupled to agree with certainty, while if there is a disagreement at z we continue recursively starting from the disagreement at z. We therefore conclude that the probability of disagreement at x in the resulting coupling of $\mu_{B_{x,\ell}}^{\tau^{w,s_1}}$ and $\mu_{B_{x,\ell}}^{\tau^{w,s_2}}$ is γ^ℓ, as required. □

Proof of Lemma 5.15: Fix an arbitrary ℓ. For every $z \in T_x$, all $s_1, s_2 \in \mathcal{S}$, and $0 \leq i \leq \ell$, let $\mathcal{H}_{z,i}^{s_1,s_2} \equiv \mathrm{E}_\nu|\sigma - \sigma'|_{z,i}$ denote the average Hamming distance at level i below z under the coupling $\nu_z^{s_1,s_2}$ of $\mu_{T_z}^{s_1}$ and $\mu_{T_z}^{s_2}$. Notice that here we also consider the case where $s_1 = s_2$ (for which $\mathcal{H}_{z,i}^{s_1,s_2} = 0$). Now by an exponential Markov inequality, to prove the lemma it is enough to show that for all (s_1, s_2), $\mathrm{E}_\nu\left[e^{t|\sigma-\sigma'|_{x,\ell}}\right] \leq e^{2et\mathcal{H}_{x,\ell}^{s_1,s_2}} \leq e^{2et\overline{\mathcal{H}}_\ell}$ for all $t \leq (2e(\ell+1)\overline{\mathcal{H}}_\ell)^{-1} \leq 1$, where $\nu = \nu_x^{s_1,s_2}$. We thus fix t as above and let $D_{z,i}^{s_1,s_2} = \mathrm{E}_\nu\left[e^{t|\sigma-\sigma'|_{z,i}}\right]$, where $\nu \equiv \nu_z^{s_1,s_2}$. Note that $D_{z,i}^{s_1,s_2}$ can be calculated recursively as follows. The main observation is that the random variable $|\sigma - \sigma'|_{z,i}$ is the sum of the b *independent* random variables $|\sigma - \sigma'|_{y,i-1}$, where y ranges over the children of z. In turn, the random variable

$e^{t|\sigma-\sigma'|_{y,i-1}}$ takes the value $D_{y,i-1}^{s_1',s_2'}$ with probability $\nu_z^{s_1,s_2}(\sigma_y = s_1', \sigma_y' = s_2')$ for every pair (s_1', s_2').

We wish to show that, for t in the above range, $D_{x,\ell}^{s_1,s_2} \le 1 + 2et\mathcal{H}_{x,\ell}^{s_1,s_2} \le e^{2et\mathcal{H}_{x,\ell}^{s_1,s_2}}$ for all (s_1, s_2). In fact, we show by induction on i that $D_{z,i}^{s_1,s_2} \le 1 + 2t[\frac{\ell+1}{\ell}]^i \mathcal{H}_{z,i}^{s_1,s_2}$ for every $z \in T_x$, all (s_1, s_2), and every $0 \le i \le \ell$. For the base case $i = 0$, notice that if $s_1 = s_2$ then $\mathcal{H}_{z,0}^{s_1,s_2} = 0$ and $D_{z,0}^{s_1,s_2} = 1$ because $|\sigma - \sigma'|_{z,0} = 0$ with certainty in this case. If $s_1 \neq s_2$ then $|\sigma - \sigma'|_{z,0} = 1$ with certainty, and therefore $\mathcal{H}_{z,0}^{s_1,s_2} = 1$ and $D_{z,0}^{s_1,s_2} = e^t \le 1 + 2t$ for t in the given range. We now assume the claim for i, every $z \in T_x$ and all pairs (s_1, s_2), and show for $i+1$, an arbitrary site z, and an arbitrary pair of spins (s_1, s_2). Let $y \prec z$ stand for "y is a child of z." Then as observed above,

$$\begin{aligned}
D_{z,i+1}^{s_1,s_2} &= \prod_{y \prec z} \left[\sum_{s_1',s_2'} \nu_z^{s_1,s_2}(\sigma_y = s_1', \sigma_y' = s_2') D_{y,i}^{s_1',s_2'} \right] \\
&\le \prod_{y \prec z} \left[1 + 2t \left[\frac{\ell+1}{\ell} \right]^i \sum_{s_1',s_2'} \nu_z^{s_1,s_2}(\sigma_y = s_1', \sigma_y' = s_2') \mathcal{H}_{y,i}^{s_1',s_2'} \right] \\
&\le \exp\left(2t \left[\frac{\ell+1}{\ell} \right]^i \sum_{y,s_1',s_2'} \nu_z^{s_1,s_2}(\sigma_y = s_1', \sigma_y' = s_2') \mathcal{H}_{y,i}^{s_1',s_2'} \right) \\
&= \exp\left(2t \left[\frac{\ell+1}{\ell} \right]^i \mathcal{H}_{z,i+1}^{s_1,s_2} \right) \\
&\le 1 + \left(\frac{\ell+1}{\ell} \right) 2t \left[\frac{\ell+1}{\ell} \right]^i \mathcal{H}_{z,i+1}^{s_1,s_2} ,
\end{aligned}$$

where for the first inequality we used the induction hypothesis, and for the last inequality we used the fact that $2t[\frac{\ell+1}{\ell}]^i \mathcal{H}_{z,i+1}^{s_1,s_2} \le \frac{1}{\ell+1}$ for all $0 \le i \le \ell$ and t in the given range. □

5.2.5 Relaxing Theorem 5.6: Variants of κ and γ

For some of our applications in Section 5.3, we will require two minor but useful generalizations of the framework for establishing bounded c_{gap} and c_{sob} that we described above. Both generalizations stem from the observation that the role of the definitions of κ and γ is to obtain the bounds on disagreement percolation stated in Claim 5.5. In fact in Theorem 5.6, we can replace κ and γ by any two values κ' and γ' for which the upper bounds in parts (i) and (ii) of Claim 5.5 are $O((\kappa' b)^\ell)$ and $O(\gamma'^\ell)$ respectively (where the $O(\cdot)$ hides

constants independent of ℓ). The arguments leading to Theorem 5.6 are all easily seen to hold in this slightly looser setting.

Our first generalization (which will be particularly useful for "non-attractive" systems) is to consider two levels of the tree at a time, rather than a single level as in Definition 5.4. Accordingly, define

$$\kappa_2 = \sup_T \max_{\substack{z,w \prec z, \\ s_1,s_2,s_1',s_2'}} \sqrt{\|\mu_{T_z}^{s_1} - \mu_{T_z}^{s_2}\|_z \cdot \|\mu_{T_w}^{s_1'} - \mu_{T_w}^{s_2'}\|_w} \,, \tag{5.18}$$

where we recall that the boundary condition at the bottom of the tree is given by the global boundary configuration η. In fact, we may restrict the maximization to sites z of even (or odd) height. With this definition, it is easy to see that the upper bound in Claim 5.5(i) for the probability of disagreement percolating down the tree can be replaced by $(\kappa_2)^{2(\ell/2-1)} b^\ell = O((\kappa_2 b)^\ell)$. We therefore get the following generalization of Theorem 5.6:

Theorem 5.6′ *In the setting of Theorem 5.6 the following hold:*

(i) *If* $\gamma \kappa_2 b < 1$ *then* $c_{\rm gap}$ *is bounded for* η.

(ii) *If* $\max\{\gamma\kappa_2 b, \gamma\} < 1$ *then* $c_{\rm sob}$ *is bounded for* η. □

Our second generalization exploits the fact that, when deriving the bound on upward percolation in part (ii) of Claim 5.5, it is enough to control the probability of a disagreement percolating upwards one level from y to z only when z is sufficiently far from the boundary and the root of $B_{x,\ell}$. For this purpose, let $\widehat{\gamma}$ be defined in the same way as γ, but with the maximization restricted to sets A that include the full subtree of depth d rooted at z under the orientation in which y is the parent of z; here d is an implicit parameter whose value may change from model to model, but will in each case be a constant independent of the size of T. Now, recall the recursive construction of the coupling in the proof of part (ii) of Claim 5.5, and the subset A used in the inductive step for choosing a coupled spin at z given a disagreement at y (see Figure 5.2). It is easy to verify that, if z is at distance at least d from the bottom and top boundaries of $B_{x,\ell}$, then the subset A includes the full subtree of depth d rooted at z under the orientation in which y is the parent of z. Therefore for such z, the probability of disagreement percolating one level upwards to z is bounded by $\widehat{\gamma}$. Thus, it is easy to modify the proof of Claim 5.5 so that the factor γ^ℓ in part (ii) is replaced by $\widehat{\gamma}^{\ell-2d} = O(\widehat{\gamma}^\ell)$.

In a similar manner, we define $\widehat{\kappa}$ as before but with the maximization restricted to z that are at distance at least d from the bottom boundary of T. Whenever we use $\widehat{\gamma}$, we will also use $\widehat{\kappa}$ with the same value of d so that we still have $\widehat{\kappa} \leq \widehat{\gamma}$. This leads to our second generalization of Theorem 5.6:

Theorem 5.6″ *In the setting of Theorem 5.6 the following hold:*

(i) *If* $\widehat{\gamma}\,\widehat{\kappa} b < 1$ *then* c_{gap} *is bounded for* η.

(ii) *If* $\max\{\widehat{\gamma}\,\widehat{\kappa} b, \widehat{\gamma}\} < 1$ *then* c_{sob} *is bounded for* η. □

5.3 Applications

In this section, we apply the framework developed in Section 5.2 for establishing bounded c_{sob} to a wide range of models and regimes of values of their parameters. In particular, for all the scenarios of interest listed below, we calculate κ and γ (or their variants given in Section 5.2.5), and show that $\max\{\gamma\kappa b, \gamma\} < 1$. By Theorem 5.6 (ii) (or, in case we use the variants, Theorem 5.6′ or 5.6″), this immediately yields bounded c_{sob}. (Of course, the weaker fact that c_{gap} is bounded also follows.)

The models we discuss are (in order of appearance): the Ising model, the hard-core model (independent sets), general two-spin models, colorings, and the ferromagnetic Potts model. The scenarios we discuss include cases in which bounded c_{sob} is established for specific boundary conditions (where c_{sob} is not bounded for other boundary conditions), as well as cases for which we prove bounded c_{sob} uniformly in the boundary condition. The detailed scenarios for each model are given at the beginning of the section dedicated to the model, where an overview of previous results is also given.

5.3.1 Ising model

An overview of the Ising model on trees was already given in Section 5.1.2, and our detailed results were stated in Theorems 5.2 and 5.3. For convenience, we restate those theorems here.

Theorem 5.2 *For any fixed b, the Glauber dynamics on the n-vertex b-ary tree with $(+)$-boundary condition has bounded c_{sob}, and therefore $O(m \log n)$ mixing time, at all inverse temperatures $\beta < \infty$ and all external fields h.*

Theorem 5.3 *For any fixed b, the Glauber dynamics on the n-vertex b-ary tree with arbitrary boundary conditions has bounded c_{sob}, and therefore $O(m \log n)$ mixing time, both (i) at all inverse temperatures $\beta < \beta_1$ and all external fields h; and (ii) at all inverse temperatures $\beta < \infty$ and all external fields $|h| > h_c(\beta)$.*

The two theorems follow from the bounds on κ and γ given below. Recall that $\gamma \equiv \gamma(\{\mu_T^\eta\})$ depends only on the potentials of the system, while $\kappa \equiv \kappa(\{\mu_T^\eta\})$ may also depend on the boundary condition η.

Theorem 5.19 *Consider the Ising model at inverse temperature β and external field h. Then:*

(i) *for all (β, h), $\gamma \leq \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}}$;*

(ii) *if (β, h) are such that the Gibbs measure is unique (i.e., $\beta \leq \beta_0$ or $|h| > h_c(\beta)$) then for every $\varepsilon > 0$ there exists a large enough d such that $\widehat{\gamma} \leq \frac{1}{b} + \varepsilon$, where d is the implicit constant in $\widehat{\gamma}$;*

(iii) *for η the all-$(+)$ boundary configuration, if (β, h) are such that the Gibbs measure is not unique (i.e., $\beta > \beta_0$ and $|h| \leq h_c(\beta)$) then $\kappa \equiv \kappa(\{\mu_T^\eta\}) \leq \frac{1}{b}$.*

Before proving Theorem 5.19, we first explain how to deduce Theorems 5.2 and 5.3 from it. Recall that following our general framework in Theorem 5.6 (ii), we wish to establish that $\max\{\gamma\kappa b, \gamma\} < 1$ in the scenarios of Theorems 5.2 and 5.3. We start with the scenario of arbitrary boundary conditions and high or intermediate temperature or large external field, as in Theorem 5.3. Recall that $\kappa \leq \gamma$ for all boundary conditions. Thus, $\max\{\gamma\kappa b, 1\} < 1$ for all boundary conditions provided that $\gamma < \frac{1}{\sqrt{b}}$. However, using part (i) of Theorem 5.19, we see that this is the case for all $\beta < \frac{1}{2}\frac{\ln(\sqrt{b}+1)}{\ln(\sqrt{b}-1)}$, i.e., for all $\beta < \beta_1$, completing the proof of Theorem 5.3 part (i). Part (ii) of Theorem 5.3 follows immediately from part (ii) of Theorem 5.19 by applying Theorem 5.6″, once we recall that $\widehat{\kappa} \leq \widehat{\gamma}$ for all boundary conditions. We go on to the scenario of the $(+)$-boundary condition, as in Theorem 5.2. Notice that the regime in which the Gibbs measure is unique is covered (for arbitrary boundary conditions) by Theorem 5.3. For the regime in which the Gibbs measure

is not unique, Theorem 5.2 follows immediately from part (iii) of Theorem 5.19, together with the fact that $\gamma < 1$ for all $\beta < \infty$, which is apparent from part (i) of the same theorem.

Proof of Theorem 5.19: To bound κ and γ, we need to bound a quantity of the form $\|\mu_A^{\tau^{y,+}} - \mu_A^{\tau^{y,-}}\|_z$, where $y \in \partial A$ and $z \in A$ is a neighbor of y. The key observation is that this quantity can be expressed very cleanly in terms of the "magnetization" at z, i.e., the ratio of probabilities of a $(-)$-spin and a $(+)$-spin at z. It will actually be convenient to work with the magnetization *without* the influence of the neighbor y: thus we let $\mu_A^{\tau^{y,*}}$ denote the Gibbs distribution with boundary condition τ, except that the spin at y is free (or equivalently, the edge connecting z to y is erased). We then have:

Proposition 5.20 *For any subset $A \subseteq T$, any boundary configuration τ, any site $y \in \partial A$ and any neighbor $z \in A$ of y, we have*

$$\|\mu_A^{\tau^{y,+}} - \mu_A^{\tau^{y,-}}\|_z = K_\beta(R),$$

where $R = \frac{\mu_A^{\tau^{y,*}}(\sigma_z=-)}{\mu_A^{\tau^{y,*}}(\sigma_z=+)}$ and the function K_β is defined by

$$K_\beta(a) = \frac{1}{e^{-2\beta}a + 1} - \frac{1}{e^{2\beta}a + 1}.$$

Proof: First, w.l.o.g. we may assume that the edge between y and z is the only one connecting y to A; this is because a tree has no cycles, so once the spin at y is fixed A decomposes into disjoint components that are independent. We abbreviate $\mu_A^{\tau^{y,+}}$, $\mu_A^{\tau^{y,-}}$ and $\mu_A^{\tau^{y,*}}$ to μ_A^+, μ_A^- and μ_A^* respectively. Thus $\|\mu_A^{\tau^{y,+}} - \mu_A^{\tau^{y,-}}\|_z = |\mu_A^+(\sigma_z = +) - \mu_A^-(\sigma_z = +)|$, and $R = \frac{\mu_A^*(\sigma_z=-)}{\mu_A^*(\sigma_z=+)}$. We write R^+ for $\frac{\mu_A^+(\sigma_z=-)}{\mu_A^+(\sigma_z=+)}$ and R^- for $\frac{\mu_A^-(\sigma_z=-)}{\mu_A^-(\sigma_z=+)}$. Since the only influence of y on A is through z, we have $R^+ = e^{-2\beta}R$ and $R^- = e^{2\beta}R$. The proposition now follows once we notice that, by definition of R^+ and R^-, $\mu_A^+(\sigma_z = +) = \frac{1}{R^+ + 1}$ and $\mu_A^-(\sigma_z = +) = \frac{1}{R^- + 1}$. □

Now it is easy to check that $K_\beta(a)$ is an increasing function in the interval $[0, 1]$, decreasing in the interval $[1, \infty]$, and is maximized at $a = 1$. Therefore, we can always bound κ and γ from above by $K_\beta(1) = \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}}$. Indeed, for γ we must make do with this crude bound because we cannot rule out the possibility that $R = 1$ when the subset A and the boundary configuration τ are arbitrary, and this completes the proof of part (i) of Theorem 5.19. However, for $\widehat{\gamma}$ and κ we only have to consider restricted scenarios (either because A includes a deep enough full subtree in the case of $\widehat{\gamma}$, or because the boundary

condition is specific in the case of κ), and as we shall see below, we get better bounds for these quantities by calculating the magnetization R in the relevant scenarios.

Before giving the details of the calculation of the magnetization, we make the following two remarks regarding $\widehat{\gamma}$. Recall that, in order to bound $\widehat{\gamma}$, we need to consider an arbitrary subset A that includes a deep enough subtree B rooted at z and an arbitrary boundary condition outside A, i.e., we wish to calculate the magnetization R at z in this setting. Now, notice that the Gibbs distribution on A is a convex combination of μ_B^σ as σ varies. Thus, if we establish that the magnetization R at z for the subtree B with an arbitrary boundary condition is at least a_1 and at most a_2, then this immediatly implies the same for the subset A. The second point we wish to make is that the Ising model is *monotone*, i.e., by changing spins on the boundary from $(+)$ to $(-)$ the magnetization $R = \frac{\Pr(\sigma_z=-)}{\Pr(\sigma_z=+)}$ can only increase. Therefore, in order to establish upper and lower bounds on the magnetization for arbitrary boundary conditions, it is enough to bound the magnetization for the all-$(+)$ and all-$(-)$ configurations respectively. Therefore, from here onwards we concentrate on calculating the magnetization at the root of full subtrees when the boundary condition at the bottom of the subtree is either all-$(+)$ or all-$(-)$. Notice also that a full subtree with the all-$(+)$ boundary is what we need to consider for bounding κ in part (iii) of Theorem 5.19.

A recursive calculation of the magnetization at the root of full subtrees

Fix a boundary configuation τ, and for a site z with parent y let $R_{z,\ell} = \frac{p(\sigma_z=-)}{p(\sigma_z=+)}$, where $p(\cdot) = \mu^{\tau^{y,*}}_{B_{z,\ell}}(\cdot)$. (If z is the root of $\mathbb{T}^b$ then $p(\cdot) = \mu^{\tau}_{B_{z,\ell}}(\cdot)$.)

We now describe a recursive calculation of the magnetization $R_{z,\ell}$, the details of which (up to change of variables) can be found in [Bax82] or [BRSSZ01]. Recall that $x \prec z$ denotes that x is a child of z. A simple direct calculation gives that $R_{z,\ell} = e^{-2\beta h}\prod_{x\prec z} F(R_{x,\ell-1})$, where $F(a) \equiv F_\beta(a) = \frac{a+e^{-2\beta}}{e^{-2\beta}a+1}$. In particular, if τ is the all-$(+)$ configuration (i.e., $R_{x,0} = 0$ for all x) then $R_{z,1} = e^{-2\beta h}[F(0)]^b$. We let $F(\infty) \equiv e^{2\beta}$ and notice that indeed, if τ is the all-$(-)$ configuration (i.e., $R_{x,0} = \infty$ for all x), then $R_{z,1} = e^{-2\beta h}[F(\infty)]^b$. From here onwards we limit our attention to the all-$(+)$ and all-$(-)$ boundary conditions. We thus define

$$J(a) \equiv J_{\beta,h}(a) = e^{-2\beta h}[F(a)]^b \tag{5.19}$$

and observe that $R_{z,\ell}$ equals $J^{(\ell)}(0)$ and $J^{(\ell)}(\infty)$ for τ all-$(+)$ and all-$(-)$ respectively, where $J^{(\ell)}$ stands for the ℓ-fold composition of J.

Now that we have expressed the magnetization in terms of the function J, our next step is relating J to the function K_β in Proposition 5.20, which expresses the total variation distance in terms of the magnetization. To begin, let us describe some properties of J that we shall use (refer to Fig. 5.3): J is continuous and increasing on $[0,\infty)$, with $J(0) = e^{-2\beta(h+b)} > 0$ and $\sup_a J(a) = e^{-2\beta(h-b)} < \infty$. This immediately implies that J has at least one fixed point in $[0,\infty)$. In fact, whether J has one or more fixed points corresponds exactly to whether the Gibbs measure (for the same values of β and h) is unique or not. This is because the Gibbs measure is unique if and only if the magnetization at the root of the tree of depth ℓ converges with ℓ to the same value conditioned on the all-$(+)$ and all-$(-)$ boundary configurations respectively. We denote by a_0 the least fixed point of J. Since a_0 is the least fixed point and $J(0) > 0$ then clearly $J'(a_0) \leq 1$, where $J'(a) \equiv \frac{\partial J(a)}{\partial a}$ is the derivative of J. In particular, when the Gibbs measure is unique, the derivative at the unique fixed point a_0 is ≤ 1. We also observe that J has a single point of inflection a_*, i.e., the derivative J' is monotonically increasing on $[0, a_*)$ and decreasing on $[a_*, \infty)$ for some $a_* \in \mathbb{R}^+$. (This follows from the fact that the equation $J''(a) = 0$ has a unique solution, as can be verified by straightforward calculus). Therefore, J can have at most three fixed points, and furthermore, if it has two or three fixed points, then necessarily $a_* \geq a_0$, i.e., $J'(a) \leq J'(a_0) \leq 1$ for $a \in [0, a_0]$.

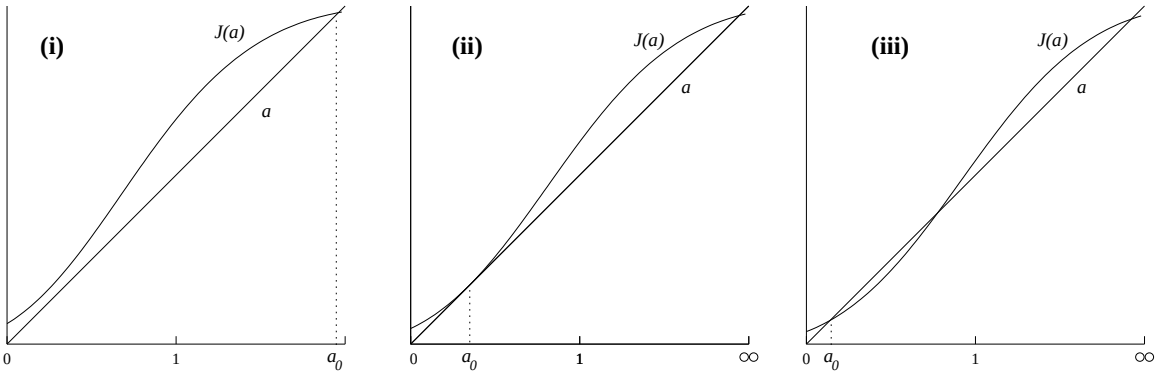

Figure 5.3: Curve of the function $J(a)$, used in the proof of Theorem 5.19, for $\beta > \beta_0$ and various values of the external field h. (i) $h < -h_c$; (ii) $h = -h_c(\beta)$; (iii) $h_c(\beta) > h > -h_c(\beta)$. The point a_0 is the smallest fixed point of J.

The relevance of the derivative J' and the fixed point a_0 to the discussion here is made clear by the following lemma.

Lemma 5.21 *For every* $a \in \mathbb{R}^+$, $K_\beta(a) = \frac{1}{b} \cdot \frac{a}{J(a)} \cdot J'(a)$.

Proof: From the definitions of J and F we have:

$$\begin{aligned} J'(a) &= e^{-2\beta h} \cdot b \cdot [F(a)]^{b-1} F'(a) \\ &= b \cdot J(a) \cdot \frac{F'(a)}{F(a)} \\ &= b \cdot \frac{J(a)}{a} \cdot a\Big[\frac{1-e^{-4\beta}}{(a+e^{-2\beta})(e^{-2\beta}a+1)}\Big] \\ &= b \cdot \frac{J(a)}{a} \cdot K_\beta(a). \qquad \square \end{aligned}$$

We are now ready to complete the proof of the bounds on $\widehat{\gamma}$ and κ given in parts (ii) and (iii) of Theorem 5.19.

Bounding $\widehat{\gamma}$ when the Gibbs measure is unique

When the Gibbs measure is unique, the magnetization $R_{z,\ell}$ converges with ℓ to the unique fixed point a_0 of J, for which $J'(a_0) \leq 1$, and thus $K_\beta(a_0) \leq \frac{1}{b}$ by Lemma 5.21. We now observe that since a_0 is the unique fixed point, for every $\varepsilon' > 0$ there exists a large enough d such that $R_{z,d} \geq a_0 - \varepsilon'$ for the all-$(+)$ boundary condition, and $R_{z,d} \leq a_0 + \varepsilon'$ for the all-$(-)$ boundary condition. As explained above, this means that for any subset A that includes the full subtree of depth d rooted at z, and with arbitrary boundary condition outside A, the relevant magnetization $R_z \in [a_0 - \varepsilon', a_0 + \varepsilon']$. (From here onwards we write R_z for the magnetization R at z as defined in Proposition 5.20, where the subset A, the boundary condition τ, and the neighbor y of z are clear from the context.) Now, since $K_\beta(a)$ is continuous in a, we deduce that that $K_\beta(R_z) \leq \frac{1}{b} + \varepsilon$ for some ε that depends on ε'. In particular, when the Gibbs measure is unique, for every $\varepsilon > 0$ there exists a large enough d such that $\widehat{\gamma} \leq \frac{1}{b} + \varepsilon$. This concludes the proof of part (ii) of Theorem 5.19.

Bounding κ for the $(+)$-boundary condition when the Gibbs measure is not unique

We now assume that η (the global boundary configuration) is all-$(+)$ and consider (β, h) such that the Gibbs measure is not unique, i.e., $\beta > \beta_0$ and $|h| \leq h_c(\beta)$. As we shall see below, the property of this regime that we use here is that J has at least two fixed points, and therefore $J'(a) \leq 1$ for $a \in [0, a_0]$.

To calculate κ, we need to bound the variation distance $\|\mu^+_{T_z} - \mu^-_{T_z}\|_z$, which by Proposition 5.20 is equal to $K_\beta(R_z)$, where $R_z = \frac{\mu^*_{T_z}(\sigma_z=-)}{\mu^*_{T_z}(\sigma_z=+)}$ and $\mu^*_{T_z}$ is the Gibbs distribution over the subtree T_z when it is disconnected from the rest of T and the spins on its

bottom boundary agree with η. Now, since η is all-$(+)$, then $R_z = J^{(\ell)}(0)$, where ℓ is the distance of z from the bottom boundary of T. We thus have $\kappa = \sup_T \max_{z \in T} K_\beta(R_z) = \sup_{\ell \geq 1} K_\beta(J^{(\ell)}(0))$.

Since J is monotonically increasing and a_0 is the least fixed point of J, then clearly $J^{(\ell)}(0)$ converges to a_0 from below, i.e., $J^{(\ell)}(0) \leq a_0$ for all ℓ. Now by applying Lemma 5.21, since $J'(a) \leq 1$ for $a \in [0, a_0]$ in the non-uniqueness regime, and since $J(a) \geq a$ for the same a, $K_\beta(J^{(\ell)}(0)) = \frac{1}{b} \cdot \frac{J^{(\ell)}(0)}{J(J^{(\ell)}(0))} \cdot J'(J^{(\ell)}(0)) \leq \frac{1}{b}$ for all ℓ. This completes the proof of part (iii) of Theorem 5.19. □

Remark: We note that in fact $\kappa \leq \frac{1}{b}$ for $\beta \leq \beta_0$ (and arbitrary h) as well. This follows from the fact that $\gamma \leq \frac{e^\beta - e^{-\beta}}{e^\beta + e^{-\beta}} \leq \frac{1}{b}$ in this regime. We also note that $\widehat{\kappa} \leq \frac{1}{b} + \varepsilon$ throughout the uniqueness regime, as is aparent from part (i) of the above theorem. Indeed, the only obstacle to proving $\kappa \leq \frac{1}{b}$ for the all-$(+)$ boundary in all (β, h) is that for $\beta < \beta_0$ and $h < -h_c(\beta)$, the derivative $J'(J^{(\ell)}(0)) > 1$ for some ℓ (see Figure 5.3). Notice, however, that this derivative converges with ℓ to a value not larger than 1. (This was used for proving $\widehat{\gamma} \leq \frac{1}{b} + \varepsilon$ in this regime.)

5.3.2 The hard-core model

We now move on to consider the hard-core model (defined in Example 2.2) on regular trees. We first review the phase diagram of this model. It is well known that the hard-core model undergoes a phase transition at a critical activity $\lambda = \lambda_0 = \frac{b^b}{(b-1)^{b+1}}$ (see, e.g., [Spi75, Kel85]). For $\lambda \leq \lambda_0$ there is a unique Gibbs measure independent of the boundary condition, while for $\lambda > \lambda_0$ there are (at least) two distinct phases, corresponding to the "odd" and "even" boundary conditions respectively. The even (odd) boundary condition is obtained by making the leaves of the tree all occupied if the depth is even (odd), and all unoccupied otherwise. For $\lambda > \lambda_0$, the probability of occupation of the root in the infinite-volume Gibbs measure differs for odd and even boundary conditions.

The Glauber dynamics for the hard-core model on trees is known to have mixing time polynomial in n at all activities $\lambda > 0$ with arbitrary boundaries [JSTV02]. Moreover, a rather general result of Luby and Vigoda [LV99, Vig01] ensures a mixing time of $O(m \log n)$ when $\lambda < \frac{2}{b-1}$, with arbitrary boundaries. This latter result actually holds for any graph G of maximum degree $b + 1$.

Our results for the hard-core model mirror those stated above for the Ising model. First, we show that c_{sob} is bounded (and therefore that the mixing time is $O(m \log n)$) for

all activities $\lambda \leq \lambda_0$ (and indeed beyond), with *arbitrary* boundary conditions. Second, for the *even* (or odd) boundary condition, we get the same result for *all* activities λ:

Theorem 5.22 *For the hard-core model on the regular b-ary tree, c_{sob} is bounded in both of the following situations:*

(i) *the boundary condition is arbitrary, and $\lambda \leq \max\left\{\lambda_0, \frac{1}{\sqrt{b}-1}\right\}$;*

(ii) *the boundary condition is even (or odd), and $\lambda > 0$ is arbitrary.*

Part (ii) of this theorem is analogous to our earlier result (Theorem 5.2) that c_{sob} is bounded for the Ising model with the $(+)$-boundary at all temperatures. This is in line with the intuition that the even boundary eliminates the only bottleneck in the dynamics. Part (i) identifies a region in which the mixing time is insensitive to the boundary condition. We would expect this to hold throughout the low-activity region $\lambda \leq \lambda_0$, and indeed, by analogy with the Ising model, also in some intermediate region beyond this. Our bound in part (i) confirms this behavior: note that the quantity $\frac{1}{\sqrt{b}-1}$ exceeds λ_0 for all $b \geq 5$, and indeed for large b it grows as $\Theta(\frac{1}{\sqrt{b}})$ compared to the $\Theta(\frac{1}{b})$ growth of λ_0. Thus for $b \geq 5$ we establish bounded c_{sob} (and $O(m \log n)$ mixing time) in a region above the critical value λ_0. To the best of our knowledge this is the first such result. (Note that the result of [LV99, Vig01] mentioned earlier establishes $O(m \log n)$ mixing time only for $\lambda < \frac{2}{b-1}$, which is less than λ_0 for all b and so does not even cover the whole uniqueness region.)

Following the standard theme in this section, we appeal to our general framework in Theorem 5.6 and its variants to deduce Theorem 5.22 from:

Theorem 5.23 *For the independent sets model with activity parameter λ:*

(i) $\widehat{\gamma} < \frac{\lambda}{1+\lambda}$ *(for $d = 3$, where d is the implicit parameter in $\widehat{\gamma}$);*

(ii) *if the Gibbs measure is unique (i.e., $\lambda \leq \lambda_0$), then for every $\varepsilon > 0$ there exists a large enough d such that $\widehat{\gamma} \leq \frac{1}{b} + \varepsilon$;*

(iii) *for η the 0-boundary condition (i.e., all sites are unoccupied), if the Gibbs measure is not unique (i.e., $\lambda > \lambda_0$) then $\kappa_2 \equiv \kappa_2(\{\mu_T^\eta\}) \leq \frac{1}{b}$.*

Recall that $\widehat{\kappa} \leq \widehat{\gamma}$, so from part (i) of this theorem we conclude that $\widehat{\gamma}\,\widehat{\kappa} b < 1$ when $(\frac{\lambda}{1+\lambda})^2 \leq \frac{1}{b}$, i.e., when $\lambda \leq \frac{1}{\sqrt{b}-1}$, and also whenever $\lambda \leq \lambda_0$ by part (ii). Since also $\widehat{\gamma} < 1$ for all

finite λ, part (i) of Theorem 5.22 follows using Theorem 5.6″. This also dispenses with part (ii) of Theorem 5.22 in the uniqueness regime. Part (ii) in the non-uniqueness regime follows immediately from part (iii) of Theorem 5.23, using Theorem 5.6′ and the fact that $\gamma < 1$. (Note that analyzing the 0-boundary for all depths of T handles both odd and even boundary conditions.)

Remark: Recall that by establishing $\widehat{\gamma}\widehat{\kappa}b < 1$ for all boundary conditions in the regime $\lambda \leq \frac{1}{\sqrt{b}-1}$, we in fact establish $\mathrm{VM}(\ell, \exp(-\Theta(\ell)))$ for all boundary conditions, and in particular, that correlations decay with distance under any boundary condition for this λ. This implies that any Gibbs measure that is the limit of finite Gibbs distributions for some fixed boundary configuration is *extremal*, extending the regime for which this was previously known. (For more on extremality under specific boundary conditions, see [BW03, Mar03].)

Proof of Theorem 5.23: The proof uses similar ideas to those used in the proof of Theorem 5.19 for the Ising model. We start with a closer look at the variation distance we need to bound in order to bound κ and γ, i.e., $\|\mu_A^{\tau^{y,1}} - \mu_A^{\tau^{y,0}}\|_z$, for some τ, A, $z \in A$, and where $y \in \partial A$ is a neighbor of z. Now, from the definition of the hard-core model, in the first distribution the site z is unoccupied with certainty, and hence the variation distance between the two distributions at z is exactly the probability that z is occupied in the *second* distribution (where y is unoccupied, or equivalently, where the edge connecting y and z is removed). Let p_z stand for this last probability. Formally,

$$\|\mu_A^{\tau^{y,1}} - \mu_A^{\tau^{y,0}}\|_z = \mu_A^{\tau^{y,0}}(\sigma_z = 1) \equiv p_z. \tag{5.20}$$

Our main goal in the rest of this proof is to bound the probability of occupation p_z for the relevant values of λ and global boundary condition η.

We start with the easy observation that, for any subset A, any boundary configuration τ and any site $z \in A$, $\mu_A^\tau(\sigma_z = 1) \leq \frac{\lambda}{1+\lambda}$, simply because the r.h.s. is the probability of z being occupied when all its neighbors are unoccupied, and if one of its neighbors is occupied than z is unoccupied with certainty. Using (5.20), we deduce that $\gamma \leq \frac{\lambda}{1+\lambda}$. We can strengthen this to $\widehat{\gamma} < \frac{\lambda}{1+\lambda}$ by noticing that equality is achieved in the above only when all neighbors of z are unoccupied with certainty, which can happen only if all neighbors of z are in ∂A or adjacent to ∂A. So by taking $d = 3$ in the definition of $\widehat{\gamma}$ we get strict inequality, and we are done with part (i) of Theorem 5.23. Parts (ii) and (iii) follow from a recursive calculation of the probabilities of occupation p_z similar to the arguments given for the Ising model, as we now describe.

A recursive calculation of the probabilities of occupation p_z

We first note that a similar calculation to the one we describe below was done in, e.g., [Kel85]. As in the Ising model, we consider subtrees $B_{z,d}$ (disconnected from the parent y of z, or equivalently, with y unoccupied) and with a fixed boundary configuration τ. Let $R_{z,d} = \frac{p_z}{1-p_z}$ stand for the ratio of probabilities that the site z is occupied and unoccupied respectively. A simple calculation verifies that $R_{z,d} = \lambda \prod_{w \prec z} \left(\frac{1}{1+R_{w,d-1}} \right)$. For each λ, we thus define the function

$$J(a) = \lambda \left(\frac{1}{1+a} \right)^b \tag{5.21}$$

and observe that, when the boundary condition is all-0 (respectively all-1), then $R_{z,d} = J^{(d)}(0)$ (respectively $J^{(d)}(\infty)$). Notice, however, that unlike the case of the Ising model, here J is monotonically decreasing. Furthermore, since $J(0) = \lambda > 0$, J has a *unique* fixed point for every λ; we denote this fixed point by $a_0 = a_0(\lambda)$. We also note that the derivatives of J alternate signs (the odd derivatives being negative). Now uniqueness of the Gibbs measure is equivalent to the fixed point a_0 being *attractive*, i.e., the derivative $J'(a_0) \geq -1$; indeed, when $\lambda = \lambda_0$ is critical the derivative at $a_0 = a_0(\lambda_0)$ is exactly -1. The equivalence between uniqueness of the Gibbs measure and the attractiveness of a_0 is better understood by considering the function $J_2(a) \equiv J^{(2)}(a) \equiv J(J(a))$, which corresponds to jumping two levels at a time. The main observation is that, since J is monotonically decreasing, J_2 must be monotonically increasing, and thus plays a similar role to that of J in the Ising model.

Let us now further describe some properties of the function J_2 (see Fig. 5.4), which can be verified using simple calculus:

1. J_2 is continuous and increasing on $[0, \infty)$, with $J_2(0) = \lambda/(1+\lambda)^b$ and $\sup_a J_2(a) = \lambda$.

2. a_0 (the unique fixed point of J) is a fixed point of J_2.

3. If the Gibbs measure is unique (i.e., $\lambda \leq \lambda_0$) then a_0 is the unique fixed point of J_2. If there are multiple Gibbs measures (i.e., $\lambda > \lambda_0$) then J_2 has three fixed points $a_1 < a_0 < a_2$, where $J(a_1) = a_2$ and $J(a_2) = a_1$.

4. The derivative $J_2'(a) \equiv \frac{\partial J_2(a)}{\partial a}$ is continuous. If a_0 is the unique fixed point of J_2 (the Gibbs measure is unique) then $J_2'(a_0) \leq 1$. If there are three fixed points then $J_2'(a_0) > 1$, and $J_2'(a) \leq 1$ for $a \in [0, a_1]$.

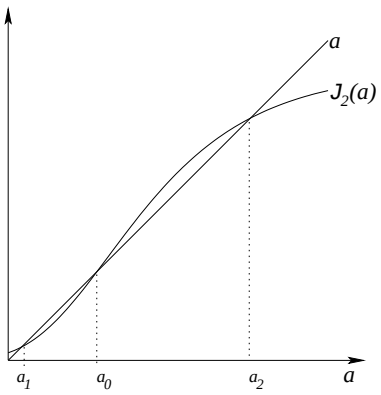

Figure 5.4: Curve of the function $J_2(a)$, used in the proof of Lemma 5.23, for $\lambda > \lambda_0$. The points a_1, a_0, a_2 are the fixed points of J_2 in increasing order.

It is now easy to see that, since $J_2'(a_0) = J'(a_0)^2$, then indeed the Gibbs measure is unique if and only if $J'(a_0) \geq -1$.

Before we go on, we wish to further clarify the connection between the uniqueness of the Gibbs measure and the uniqueness of the fixed point of J_2. First, notice that for odd-depth (respectively, even-depth) trees, the probability of occupation at the root is monotonically decreasing (respectively, increasing) in the boundary configuration. In particular, for both even and odd depths, the all-0 and all-1 boundaries are the extreme boundary configurations, i.e., for all boundary conditions the probability of occupation at the root is in the range delimited by the probabilities under the all-0 and all-1 boundary conditions. Thus, when J_2 has a unique fixed point, the probability p_z of occupation at the root of even-depth trees converges with the depth of the tree to the same value for all boundary configurations. This also means that p_z converges to the above value in odd-depth trees (uniformly in the boundary condition), because the limit for even-depth trees with the all-0 boundary is the same as the limit for odd-depth trees with the all-1 boundary, and vice versa.

With the above recursive calculation of p_z at hand, we can now complete our arguments for bounding $\widehat{\gamma}$ and κ.

Bounding $\widehat{\gamma}$ when the Gibbs measure is unique

Here it is enough to show that, when the Gibbs measure is unique, then $p_0 \leq \frac{1}{b}$, where p_0 is the value to which the probability p_z converges with the depth of the tree. This is

because, following an explanation similar to that given in the proof for the Ising model, for every $\varepsilon > 0$ there exists a large enough d such that, for any subset A that includes the full subtree of depth d, $p_z \in [p_0 - \varepsilon, p_0 + \varepsilon]$ uniformly in the boundary condition. In particular, $\widehat{\gamma} \leq p_0 + \varepsilon \leq \frac{1}{b} + \varepsilon$.

To see that $p_0 \leq \frac{1}{b}$ we observe that $p_0 = \frac{a_0}{1+a_0}$, where a_0 is the fixed point of J, and recall that $J'(a_0) \geq -1$ in the regime of λ we consider. Now note that, by a straightforward calculation,

$$J'(a) = -b \cdot \frac{J(a)}{1+a}. \tag{5.22}$$

Thus, since a_0 is a fixed point of J, we have $-J'(a_0) = b \cdot \frac{J(a_0)}{1+a_0} = b \cdot \frac{a_0}{1+a_0} = bp_0$. On the other hand, since $-J'(a_0) \leq 1$, we conclude that $p_0 \leq \frac{1}{b}$, as required. This completes the proof of part (ii) of Theorem 5.23.

Bounding κ for the all-0 boundary when the Gibbs measure is not unique

Here the boundary condition η is set to all-0 and $\lambda > \lambda_0$, i.e, the Gibbs measure is not unique. Once again, our aim is to calculate the probabilities of occupation p_z. Here, however, A is a maximal subtree T_z and the bottom boundary condition is η (rather than arbitrary). We start by noticing that, since η is the all-0 configuration, then $\frac{p_z}{1-p_z} \equiv R_z = J^{(\ell)}(0)$, where ℓ is the distance of z from the bottom boundary of T. Notice that since we are in the regime of non-uniqueness of the Gibbs measure, the sequence $J^{(\ell)}(0)$ does not converge with ℓ to a_0, but oscillates around it, i.e., $J^{(2\ell)}(0) = J_2^{(\ell)}(0)$ converges to a_1 while $J^{(2\ell+1)}(0) = J(J_2^{(\ell)})$ converges to a_2.

Recall that $\kappa_2^2 = \sup_T \max_{w \prec z} \|\mu^1_{T_z} - \mu^0_{T_z}\|_z \cdot \|\mu^1_{T_w} - \mu^0_{T_w}\|_w = \max_{w \prec z} p_z p_w$, and that it is enough to consider sites z whose height in T is odd. Therefore, by letting $K(R) = \frac{R}{1+R}$ (i.e., K translates the ratio $R = \frac{p}{1-p}$ to p), we see that $\kappa_2^2 = \sup_{\ell \geq 1} K[J(J_2^{(\ell)}(0))] \cdot K[J_2^{(\ell)}(0)]$. We will use the properties of J_2 in order to show that the last expression is bounded by $\frac{1}{b^2}$.

Again we show a connection between the derivative (this time of J_2) and the

relevant probabilities of occupation $K(J(R)) \cdot K(R)$. Using (5.22) to calculate $J_2'(a)$ gives:

$$\begin{aligned} J_2'(a) &= J'(J(a)) \cdot J'(a) \\ &= b\left(\frac{J(J(a))}{1+J(a)}\right) \cdot b\left(\frac{J(a)}{1+a}\right) \\ &= b^2 \cdot \frac{J_2(a)}{a} \cdot \frac{J(a)}{1+J(a)} \cdot \frac{a}{1+a} \\ &= b^2 \cdot \frac{J_2(a)}{a} \cdot K(J(a)) \cdot K(a). \end{aligned}$$

We now recall that $J_2^{(\ell)}(0)$ converges to a_1 from below, and that $J_2(a) \geq a$ and $J_2'(a) < 1$ for all $a \in [0, a_1]$, to conclude that for all ℓ,

$$K[J(J_2^{(\ell)}(0))] \cdot K[J_2^{(\ell)}(0)] \;=\; \frac{1}{b^2} \cdot \frac{J_2^{(\ell)}(0)}{J_2(J_2^{(\ell)}(0))} \cdot J_2'(J_2^{(\ell)}(0))) \;\leq\; \frac{1}{b^2},$$

as required. This completes the proof of part (iii) of Theorem 5.23. □

5.3.3 General two-spin systems

It is not too difficult to see that most of our analysis of the hard-core model was based on the same high-level ideas as the analysis of the Ising model. Indeed, in this subsection we show that these ideas are part of a general theory that holds for any spin system for which the spin space $\mathcal{S}$ consists of two values, and for which the pair- and self-potentials are uniform in the edges and sites of the tree, respectively.

Here we will consider the following specification of two-spin systems with uniform potentials. First, w.l.o.g. we can assume that $\mathcal{S} = \{-, +\}$. Furthermore, we can assume that the value of the pair potential $U_{\{x,y\}}(-,+) < \infty$ since otherwise the system is trivial with only two feasible configurations (all-$(+)$ and all-$(-)$ respectively). Now, since the Gibbs distribution remains unaffected by adding a uniform constant value to the potential, we can assume w.l.o.g. that $U_{\{x,y\}}(-,+) = 0$, and hence that the pair potential is specified by the two values $U_{\{x,y\}}(-,-)$ and $U_{\{x,y\}}(+,+)$. We let $\lambda_{(-)} = \exp(-U_{\{x,y\}}(-,-))$ and $\lambda_{(+)} = \exp(-U_{x,y}(+,+))$. In a similar manner, we can assume w.l.o.g. that the self potential $U_x(+) = 0$, and let $\lambda = \exp(-U_x(-))$. Notice that the Gibbs distribution assigns to configuration σ probability proportional to $\lambda^{\#\{-\}} \cdot (\lambda_{(-)})^{\#\{-,-\}} \cdot (\lambda_{(+)})^{\#\{+,+\}}$, where $\#\{-\}, \#\{-,-\}$ and $\#\{+,+\}$ stand for the number of sites whose spin is $(-)$, edges whose spins are $\{-,-\}$ and edges whose spins are $\{+,+\}$ respectively in σ. From here onwards we

assume that a two-spin system is specified by the three parameters $(\lambda, \lambda_{(-)}, \lambda_{(+)})$. For example, the Ising model with parameters (β, h) is given by $\lambda = e^{-2\beta h}$ and $\lambda_{(-)} = \lambda_{(+)} = e^{2\beta}$. The hard-core model with activity λ is given by λ, $\lambda_{(-)} = 0$ and $\lambda_{(+)} = 1$, where we have identified the spins 0 and 1 of the hard-core model with $(+)$ and $(-)$ respectively. Before stating our results for two-spins systems, we note that the case $\lambda_{(-)} = \lambda_{(+)} = 0$ corresponds to a trivial system with only two feasible configurations (the "odd" and "even" configurations respectively, in which the spin values alternate along the levels of the tree). We thus assume w.l.o.g. that in any given system $\lambda_{(+)} > 0$.

Our results for general two-spins systems are summarized in the following theorem:

Theorem 5.24 *For any two-spin system $(\lambda, \lambda_{(-)}, \lambda_{(+)})$ on the regular b-ary tree:*

(i) *if the Gibbs measure is unique then c_{sob} is bounded uniformly in the boundary condition;*

(ii) *c_{sob} is bounded for the all-$(+)$ (and therefore also for the all-$(-)$) boundary condition.*

As usual, this follows from bounds on κ, γ:

Theorem 5.25 *For any two-spin system $(\lambda, \lambda_{(-)}, \lambda_{(+)})$ on the regular b-ary tree:*

(i) *if the Gibbs measure is unique then, for every $\varepsilon > 0$, there exists a large enough d such that $\widehat{\gamma} \leq \frac{1}{b} + \varepsilon$;*

(ii) *for η the all-$(+)$ configuration, if the Gibbs measure is not unique then $\kappa_2 \equiv \kappa_2(\{\mu_T^\eta\}) \leq \frac{1}{b}$.*

We observe that $\gamma < 1$ for every two-spin system (because we are assuming $\lambda_{(+)} > 0$ and thus for any boundary condition, the spin at any given site is $(+)$ with positive probability), and hence Theorem 5.24 follows from Theorem 5.25 by applying either Theorem 5.6″ in case the Gibbs measure is unique, or Theorem 5.6′ for the $(+)$-boundary in case the Gibbs measure is not unique.

Proof of Theorem 5.25: The first step in the proof is relating the total variation distance at z between two distributions that differ in a single boundary site y to the "magnetization" at z. The following is a generalization of Proposition 5.20:

Proposition 5.26 *For any subset $A \subseteq T$, any boundary configuration τ, any site $y \in \partial A$ and any neighbor $z \in A$ of y, we have*

$$\|\mu_A^{\tau^{y,+}} - \mu_A^{\tau^{y,-}}\|_z = |K(R)|,$$

where $R = \frac{\mu_A^{\tau^{y,*}}(\sigma_z=-)}{\mu_A^{\tau^{y,*}}(\sigma_z=+)}$ and the function K is defined by

$$K(a) = \frac{\lambda_{(+)}}{a+\lambda_{(+)}} - \frac{1}{\lambda_{(-)}a+1}.$$

Proof: First, as was already explained in the proof of Proposition 5.20, w.l.o.g. we may assume that the edge between y and z is the only one connecting y to A. As in the previous proof, we abbreviate $\mu_A^{\tau^{y,+}}$, $\mu_A^{\tau^{y,-}}$ and $\mu_A^{\tau^{y,*}}$ to μ_A^+, μ_A^- and μ_A^* respectively. Thus $\|\mu_A^{\tau^{y,+}} - \mu_A^{\tau^{y,-}}\|_z = |\mu_A^+(\sigma_z = +) - \mu_A^-(\sigma_z = +)|$, and $R = \frac{\mu_A^*(\sigma_z=-)}{\mu_A^*(\sigma_z=+)}$. We write R^+ for $\frac{\mu_A^+(\sigma_z=-)}{\mu_A^+(\sigma_z=+)}$ and R^- for $\frac{\mu_A^-(\sigma_z=-)}{\mu_A^-(\sigma_z=+)}$. Again, since the only influence of y on A is through z, we have $R^+ = \frac{R}{\lambda_{(+)}}$ and $R^- = R\lambda_{(-)}$. The proposition now follows once we notice that, by definition of R^+ and R^-, $\mu_A^+(\sigma_z = +) = \frac{1}{R^++1}$ and $\mu_A^-(\sigma_z = +) = \frac{1}{R^-+1}$. □

The next step is generalizing the recursive calculation of the magnetization at the root of full subtrees. As we did for the Ising model, fix a boundary configuation τ, and for a site z with parent y let $R_{z,\ell} = \frac{p(\sigma_z=-)}{p(\sigma_z=+)}$, where $p(\cdot) = \mu_{B_{z,\ell}}^{\tau^{y,*}}(\cdot)$. (If z is the root of $\mathbb{T}^b$ then $p(\cdot) = \mu_{B_{z,\ell}}^{\tau}(\cdot)$.) A direct calculation (similar to that for the Ising model) gives that $R_{z,\ell} = \lambda \prod_{x \prec z} F(R_{x,\ell-1})$, where $F(a) = \frac{\lambda_{(-)}a+1}{a+\lambda_{(+)}}$. In particular, if τ is the all-$(+)$ configuration (i.e., $R_{x,0} = 0$ for all x) then $R_{z,1} = \lambda[F(0)]^b$. Again, we let $F(\infty) \equiv \lambda_{(-)}$ and notice that indeed, if τ is the all-$(-)$ configuration (i.e., $R_{x,0} = \infty$ for all x), then $R_{z,1} = \lambda[F(\infty)]^b$. As before, this motivates us to define

$$J(a) = \lambda[F(a)]^b, \tag{5.23}$$

where we notice that $R_{z,\ell}$ equals $J^{(\ell)}(0)$ and $J^{(\ell)}(\infty)$ for τ all-$(+)$ and all-$(-)$ respectively.

We now observe that the relationship established for the Ising and hard-core models between K and the derivative of J holds for general two-spin systems. In particular, a straightforward calculation verifies that $K(a) = aF'(a)/F(a)$ for all a, and therefore

$$J'(a) \;=\; bJ(a)\frac{F'(a)}{F(a)} \;=\; b \cdot \frac{J(a)}{a} \cdot K(a). \tag{5.24}$$

The proof of Theorem 5.25 will be concluded once we notice that the function J here has the same properties and relationships with the uniqueness of the Gibbs measure

as those mentioned earlier for the Ising and hard-core models. To this end, we separate the discussion into two classes of systems, where we think of the Ising and hard-core models as the representatives of each class. We say that a system is *ferromagnetic* (respectively, *antiferromagnetic*) if $\lambda_{(-)} \cdot \lambda_{(+)} > 1$ (respectively, if $\lambda_{(-)} \cdot \lambda_{(+)} < 1$). Notice that in a ferromagnetic system neighboring spins are positively correlated, i.e., the spin at site z is more likely to be $(+)$ conditioned on its neighbor being $(+)$ than conditioned on its neighbor being $(-)$. In an antiferromagnetic system, the opposite effect takes place. Indeed, if $\lambda_{(-)} \cdot \lambda_{(+)} = 1$ then the spin at z is independent of the spins of its neighbors. Notice that in the latter case the Gibbs distribution is a product distribution, and trivially $\kappa = \gamma = 0$ for all boundary conditions.

We now describe the properties of the function J, first for ferromagnetic and then for antiferromagnetic systems.

Ferromagnetic systems

For a ferromagnetic system, the function J has all the properties we described earlier in the Ising model case. Specifically:

1. J is continuous and increasing on $[0, \infty)$, with $J(0) = \lambda(1/\lambda_{(+)})^b > 0$ and $\sup_a J(a) = \lambda(\lambda_{(-)})^b < \infty$.

2. J has at least one fixed point in $[0, \infty)$. The fixed point is unique if and only if the system admits a unique Gibbs measure.

3. $J'(a_0) \leq 1$, where a_0 denotes the least fixed point of J. In particular, if the Gibbs measure is unique then the derivative at the unique fixed point is $J'(a_0) \leq 1$.

4. If a_0 is not the unique fixed point then $J'(a) \leq 1$ for $a \in [0, a_0]$. (This follows from the fact that J has a single point of inflection point a_*, i.e., the derivative J' is monotonically increasing on $[0, a_*)$ and decreasing on $[a_*, \infty)$ for some $a_* \in \mathbb{R}^+$, which also means that J has at most three fixed points.)

We note that the relationship with the uniqueness of the Gibbs measure follows from the fact that the all-$(-)$ and all-$(+)$ configurations are the *minimal* and *maximal* boundary conditions respectively, as in the case of the Ising model, i.e., for any boundary condition

the probability that the spin at z is $(+)$ is bounded below and above by the same probability under the all-$(-)$ and all-$(+)$ boundary conditions respectively.

Now, part (i) of Theorem 5.25 follows by the same argument used for the Ising model in the uniqueness regime (see the proof of Theorem 5.19) since the variation distance for the fixed-point magnetization $|K(a_0)| = \frac{1}{b}|J'(a_0)| \leq \frac{1}{b}$. For part (ii) we repeat the observation that, for η the all-$(+)$ configuration, $\kappa \equiv \kappa(\{\mu_T^\eta\}) = \sup_{\ell \geq 1} K[J^{(\ell)}(0)]$. Since $|K(a)| = \frac{1}{b}\frac{J(a)}{a}|J'(a)| \leq \frac{1}{b}$ for $a \in [0, a_0]$, this implies $\kappa \leq \frac{1}{b}$. Notice that for ferromagnetic systems we get the stronger result that κ is bounded rather than just κ_2. ($\kappa_2 \leq \kappa$ always.) This completes the proof of Theorem 5.25 for ferromagnetic systems.

Antiferromagnetic systems

For an antiferromagnetic system, the function J has all the properties described earlier for the hard-core model, and we again refer to the function $J_2(a) \equiv J(J(a))$. Specifically, the two functions have the following properties:

1. J is continuous and decreasing on $[0, \infty)$, with $0 < J(0) = \lambda(1/\lambda_{(+)})^b < \infty$ and $J(\infty) = \lambda(\lambda_{(-)})^b \geq 0$.

2. J has a unique fixed point a_0.

3. The derivatives of J alternate signs (the odd derivatives are negative).

4. J_2 is continuous and increasing on $[0, \infty)$, with $J_2(0) = J(J(0)) > J(\infty) \geq 0$ and $\sup_a J_2(a) \leq J(0) < \infty$.

5. a_0 (the unique fixed point of J) is a fixed point of J_2.

6. If the system admits a unique Gibbs measure then a_0 is the unique fixed point of J_2. If there are multiple Gibbs measures then J_2 has three fixed points $a_1 < a_0 < a_2$, where $J(a_1) = a_2$ and $J(a_2) = a_1$.

7. The derivative $J_2'(a)$ is continuous. If a_0 is the unique fixed point of J_2 (the Gibbs measure is unique) then $J_2'(a_0) \leq 1$. If there are three fixed points then $J_2'(a_0) > 1$, and $J_2'(a) < 1$ for $a \in [0, a_1]$. Consequently, the system admits a unique Gibbs measure if and only if $|J'(a_0)| = \sqrt{J_2'(a_0)} \leq 1$.

Again, the connection with the uniqueness of the Gibbs measure stems from the fact that the all-$(+)$ and all-$(-)$ are extreme configurations (though, as in the hard-core model, the direction of the monotonicity depends on the parity of the depth of the tree).

We also observe that

$$\begin{aligned} J_2'(a) &= J'(J(a)) \cdot J'(a) \\ &= b \cdot \frac{J(J(a))}{J(a)} \cdot K(J(a)) \cdot b \cdot \frac{J(a)}{a} \cdot K(a) \\ &= b^2 \cdot \frac{J_2(a)}{a} \cdot K(J(a))K(a). \end{aligned}$$

Now, part (i) of Theorem 5.25 (the uniqueness case) follows by the same argument used for the ferromagnetic case, since $|K(a_0)| = \frac{1}{b}|J'(a_0)| \leq \frac{1}{b}$ when the Gibbs measure is unique. Part (ii) of Theorem 5.25 follows from the same arguments used for the hard-core model in the non-uniqueness regime, once we notice that, as in the hard-core case, for η the all-$(+)$ configuration, $\kappa \equiv \kappa(\{\mu_T^\eta\}) = \sup_{\ell \geq 1} |K[J(J_2^{(\ell)}(0))]| \cdot |K[J_2^{(\ell)}(0)]|$. This completes the proof of Theorem 5.25 for antiferromagnetic systems. □

5.3.4 Colorings

We now move on to consider systems with more than two spin values. The first such model we consider is colorings (as defined in Example 2.4).

For colorings on the b-ary tree it is well known that, when $q \leq b + 1$, there are multiple Gibbs measures; this follows immediately from the existence of "frozen configurations," i.e., colorings in which the color of every internal vertex is forced by the colors of the leaves (see, e.g., [BW02]). Recently it has been proved that, as soon as $q \geq b + 2$, the Gibbs measure is unique [Jon02].

The sharpest result known for the Glauber dynamics on colorings is due to Vigoda [Vig00], who shows that for arbitrary boundary conditions the mixing time is $O(m \log n)$ provided $q > \frac{11}{6}(b + 1)$. This result actually holds not only for trees but for *any* n-vertex graph G of maximum degree $b + 1$. For graphs of large maximum degree and girth at least 6, this range was recently improved [DFPV04] to $q > \max\{1.489(b + 1), q_0\}$, where q_0 is an absolute constant.[4] Before we state our results for the Glauber dynamics on colorings,

[4] A recent sequence of papers [DF01, Mol02, Hay03] have reduced the required number of colors further for general graphs, under the assumption that the maximum degree is $\Omega(\log n)$. The current state of the art requires $q \geq (1 + \epsilon)(b + 1)$, for arbitrarily small $\epsilon > 0$ [HV03], but these results do not apply in our setting where the degree $b + 1$ is fixed.

notice that the dynamics is connected for $q \geq b+3$ (on any graph of maximum degree $b+1$, as was mentioned in Section 2.3.1). For $q \leq b+1$, there is at least one boundary condition η for which the dynamics for μ_T^η is not connected. For the critical value $q = b+2$, the situation is somewhat delicate: while the dynamics is connected for all boundary conditions when run on T, it is not connected for at least one boundary condition if we add a boundary site above T (i.e., if the dynamics is run on T_x for x not the root of $\mathbb{T}^b$). Furthermore, even on T the dynamics is not connected for some boundary condition, if we consider the version of $\mathbb{T}^b$ in which the root has $b+1$ children (i.e., the Bethe lattice - see footnote 1 in Section 5.1.2). Since our arguments apply equally well to these settings, the smallest q for which we can hope to establish bounded c_{sob} uniformly in the boundary condition is $q = b+3$. Indeed, we establish this for the entire regime in which the Glauber dynamics is guaranteed to be connected, i.e., for $q \geq b+3$. We note that this is the first result that establishes this fact for a non-trivial graph. (It has been conjectured that the dynamics mixes in $O(m \log n)$ time for $q \geq b+3$ on *any* graph of maximum degree $b+1$.) We also notice that, if the Glauber dynamics is replaced by the heat-bath dynamics based on flipping *edges* (i.e., where the collection of blocks is the edges of the tree), then the dynamics remains connected for $q = b+2$ and all subsets and boundary conditions. It is not too difficult to see that our results below imply bounded c_{sob} uniformly in the boundary condition for this dynamics even at the critical value $q = b+2$. Thus, we essentially establish bounded c_{sob} uniformly in the boundary condition throughout the uniqueness regime.

Theorem 5.27 *For the colorings model with q colors on the b-ary tree, c_{sob} is bounded for arbitrary boundary conditions provided $q \geq b+3$.*

Our approach for proving Theorem 5.27 is similar to that we used previously for the uniqueness regime of two-spin systems. Specifically, we show:

Theorem 5.28 *For the colorings model with q colors, if the infinite-volume Gibbs measure is unique then for every $\varepsilon > 0$ we have $\widehat{\gamma} \leq \frac{1}{q-1} + \varepsilon$ (for a suitable choice $d = d(\varepsilon)$ of the implicit constant in $\widehat{\gamma}$).*

Since in [Jon02] it was shown that the Gibbs measure is unique for all $q \geq b+2$, we conclude that for these values of q, $\widehat{\gamma} \leq \frac{1}{q-1} + \varepsilon < \frac{1}{b}$. Theorem 5.27 now follows from Theorem 5.6″ as usual.

Remarks:

- The reader may wonder why the implication $\widehat{\gamma}\widehat{\kappa}b < 1 \Rightarrow c_{\rm sob}$ fails for the critical value $q = b + 2$. Namely, a natural question to ask is: where in our framework developed in Section 5.2 did we use the assumption that the dynamics is connected? This assumption was in fact used in (5.2), i.e., when we compared the local entropy in $B_{x,\ell}$ to that at single sites, we assumed that $c_{\rm sob}$ of the Glauber dynamics in the block $B_{x,\ell}$ is bounded away from zero uniformly in the boundary condition. Notice that the latter assumption does not hold if the Glauber dynamics in $B_{x,\ell}$ is not connected for some boundary configuration. This is because $c_{\rm sob}(P) = c_{\rm gap}(P) = 0$ for any disconnected Markov chain P.

- The fact that $\widehat{\gamma} < \frac{1}{b}$ even for the critical value $q = b + 2$ is what allows us to deduce bounded $c_{\rm sob}$ for the edge dynamics mentioned above. This is because our general framework in Section 5.2 can easily be extended to the edge dynamics. (All that needs to be changed is inequality (5.2) in which local entropy in $B_{x,\ell}$ is translated to that at single sites, and the translation should be to edges instead.)

- The fact that $\widehat{\gamma} < \frac{1}{b}$ means not only that the influence of *any* boundary configuration on the spin at the root decays with the distance of the boundary from the root (as is already implied by the fact the Gibbs measure is unique), but that it decays *exponentially fast*. This fact is of independent interest, and to the best of our understanding was not obvious from the proof of uniqueness given in [Jon02].

Proof of Theorem 5.28: The idea of the proof is the following. Consider a subset A, a site $y \in \partial A$ and $z \in A$, where z is a neighbor of y. Fix a disagreement at y, i.e., set two different spins s_1 and s_2 at y. We wish to bound $\|\mu_A^{\tau^{y,s_1}} - \mu_A^{\tau^{y,s_2}}\|_z$. Suppose now that, without the condition at y, the color of z is equally likely to be any of the q possible colors (as is the case when the boundary around A is "free"). Thus, the distribution of the color of z in $\mu_A^{\tau^{y,s_1}}$ is uniform over the $q-1$ colors other than s_1, and in $\mu_A^{\tau^{y,s_2}}$ it is uniform over the colors other than s_2. We then get that $\|\mu_A^{\tau^{y,s_1}} - \mu_A^{\tau^{y,s_2}}\|_z = \frac{1}{q-1}$, because we can couple the two distributions such that spin s_2 in $\mu_A^{\tau^{y,s_1}}$ is coupled with spin s_1 in $\mu_A^{\tau^{y,s_2}}$ and the two spins at z agree otherwise. (It is easy to see that this is the optimal coupling.) In our proof, we use the hypothesis that the Gibbs measure is unique in order to approximate the Gibbs distribution over A under any boundary configuration by the free boundary case, and hence get that the variation distance is arbitrarily close (as a function of the implicit parameter d) to $\frac{1}{q-1}$.

Let us proceed with the formal proof. Recall that in order to bound $\widehat{\gamma}$, we need to consider a subset A that includes the full subtree of depth d rooted at z, and bound

the variation distance $\max_{s_1,s_2} \|\mu_A^{\tau^{y,s_1}} - \mu_A^{\tau^{y,s_2}}\|_z$ for an arbitrary boundary configuration τ, where y is the (unique) neighbor of z in ∂A.

Now it is easy to see that, for the colorings model,

$$\|\mu_A^{\tau^{y,s_1}} - \mu_A^{\tau^{y,s_2}}\|_z = \max\{\mu_A^{\tau^{y,s_1}}(\sigma_z = s_2), \mu_A^{\tau^{y,s_2}}(\sigma_z = s_1)\}.$$

This identity follows from an argument similar to that used above for the free boundary case, as we now explain. Observe that, for every color s that differs from both s_1 and s_2, $\mu_A^{\tau^{y,s_1}}(\sigma_z = s) = \frac{1-\mu_A^{\tau^{y,s_1}}(\sigma_z=s_2)}{1-\mu_A^{\tau^{y,s_2}}(\sigma_z=s_1)} \cdot \mu_A^{\tau^{y,s_2}}(\sigma_z = s)$. Hence, if $\mu_A^{\tau^{y,s_1}}(\sigma_z = s_2) \geq \mu_A^{\tau^{y,s_2}}(\sigma_z = s_1)$, then $\mu_A^{\tau^{y,s_1}}(\sigma_z = s) \leq \mu_A^{\tau^{y,s_2}}(\sigma_z = s)$ for every $s \neq s_2$, and so the event $\mathcal{E} = \{\sigma_z = s_2\}$ maximizes the expression $|\mu_A^{\tau^{y,s_1}}(\mathcal{E}) - \mu_A^{\tau^{y,s_2}}(\mathcal{E})|$ over all events $\mathcal{E}$ that only depend on σ_z.

As in the previous models we analyzed, it is now convenient to consider the distribution induced by removing the edge from z to y (i.e., with a "free" condition at y). Recall that this distribution is denoted $\mu_A^{\tau^{y,*}}$. Let $p_z(s) = \mu_A^{\tau^{y,*}}(\sigma_z = s)$, and notice that for the colorings model $\mu_A^{\tau^{y,s_1}}(\sigma_z = s_2) = \frac{p_z(s_2)}{1-p_z(s_1)}$ simply because $\mu_A^{\tau^{y,s_1}}(\,\cdot\,) = \mu_A^{\tau^{y,*}}(\cdot \,|\, \sigma_z \neq s_1)$. Thus, $\max_{s_1,s_2} \|\mu_A^{\tau^{y,s_1}} - \mu_A^{\tau^{y,s_2}}\|_z = \max_{s_1,s_2} \mu_A^{\tau^{y,s_1}}(\sigma_z = s_2) = \max_{s_1,s_2} \frac{p_z(s_2)}{1-p_z(s_1)}$.

To obtain the claimed bound on $\widehat{\gamma}$ we have to show that, for all sets A as above, and all boundary configurations τ, $\max_{s_1,s_2} \frac{p_z(s_2)}{1-p_z(s_1)} \leq \frac{1}{q-1} + \varepsilon$. It is at this point that we use the assumption that the Gibbs measure is unique. This means that, if d (the depth of the full subtree contained in A) is large enough, the distribution $p_z(\,\cdot\,)$ is arbitrarily close to the uniform distribution, regardless of the boundary configuration τ. Thus, for every $\varepsilon' > 0$, there exists a (large enough) constant d such that $p_z(s) \leq \frac{1+\varepsilon'}{q}$ for all colors s. Hence, $\max_{s_1,s_2} \frac{p_z(s_2)}{1-p_z(s_1)} \leq \frac{1+\varepsilon'}{q-1-\varepsilon'} \leq \frac{1}{q-1} + \varepsilon$ for some ε that goes to zero with d, as required. □

5.3.5 The ferromagnetic Potts model

The last model we analyze is the (ferromagnetic) Potts model (as defined in Example 2.3). Qualitatively the behavior of this model is similar to that of the Ising model, though less is known in precise quantitative terms. Again there is a phase transition at a critical $\beta = \beta_0$, which depends on b and q, so that for $\beta > \beta_0$ (and indeed for $\beta \geq \beta_0$ when $q > 2$) there are multiple phases. This value β_0 does not in general have a closed form, but it is known [Hag96] that $\beta_0 < \frac{1}{2}\ln(\frac{b+q-1}{b-1})$ for all $q > 2$. (For $q = 2$, this value is exactly β_0 for the Ising model as quoted earlier.)

Little is known about the Glauber dynamics for the Potts model on trees, beyond the facts that the mixing time is $O(m \log n)$ for arbitrary boundaries at very high temperatures (by the Dobrushin Uniqueness Condition), and is $\Omega(mn^{\epsilon})$ for some boundaries at very low temperatures (combining results in [KMP01, MP01]). Here we prove:

Theorem 5.29 *The Glauber dynamics for the Potts model on the b-ary tree has bounded $c_{\rm sob}$ in all of the following situations:*

(i) *the boundary condition is arbitrary and $\beta < \max\left\{\beta_0, \frac{1}{2}\ln(\frac{\sqrt{b}+1}{\sqrt{b}-1})\right\}$;*

(ii) *the boundary condition is constant (e.g., all sites on the boundary have spin 1) and β is arbitrary*;

(iii) *the boundary condition is free (i.e., the boundary spins are unconstrained) and $\beta < \beta_1$, where β_1 is the solution to the equation $\frac{e^{2\beta_1}-1}{e^{2\beta_1}+q-1} \cdot \frac{e^{2\beta_1}-1}{e^{2\beta_1}+1} = \frac{1}{b}$.*

Part (i) of this theorem shows that we get bounded $c_{\rm sob}$ for arbitrary boundaries throughout the uniqueness region; also, since $\frac{1}{2}\ln(\frac{\sqrt{b}+1}{\sqrt{b}-1}) \geq \frac{1}{2}\ln(\frac{b+q-1}{b-1}) > \beta_0$ when $q \leq 2(\sqrt{b}+1)$, this result extends into the multiple phase region for many combinations of b and q. Part (ii) of the theorem is an analog of our earlier result that in the Ising model $c_{\rm sob}$ is bounded for the $(+)$-boundary at all temperatures. Part (iii) is of interest for two reasons. First, since $\beta_1 > \beta_0$ always, it exhibits a natural boundary condition under which $c_{\rm sob}$ is bounded beyond the uniqueness region (but not for arbitrary β) *for all* combinations of b and q. Second, because of an intimate connection between the free boundary case and so-called "reconstruction problems" on trees [Mos02] (in which the edges are noisy channels and the goal is to reconstruct a value transmitted from the root), we obtain an alternative proof of the best known value of the noise parameter under which reconstruction is impossible [MP01] (i.e., the best known bound on the regime for which $\mathrm{VM}(\ell, \varepsilon)$ holds with ε going to zero with ℓ). As we observe later, a slight strengthening of part (iii) marginally improves on this threshold.

The following theorem sets out the relevant properties of κ and γ from which we deduce Theorem 5.29.

Theorem 5.30 *For the Potts model with q colors at inverse temperature β the following hold:*

(i) *$\gamma \leq \frac{e^{2\beta}-1}{e^{2\beta}+1} - \delta$, where $\delta = \delta(b, q, \beta) \geq 0$ with equality if and only if $q = 2$. Furthermore, $\delta(b, q, \beta)$ is increasing in q and decreasing in b and β. [The exact definition of $\delta(b, q, \beta)$ is rather involved and given in the proof below.]*

(ii) *If the Gibbs measure is unique (i.e., $\beta < \beta_0$) then $\widehat{\gamma} \leq \frac{e^{2\beta}-1}{e^{2\beta}+q-1} + \varepsilon < \frac{1}{b}$ (for a small enough ε that depends on the choice of d, the implicit constant in $\widehat{\gamma}$).*

(iii) *If the Gibbs measure is not unique and the boundary condition η is constant then $\kappa \equiv \kappa(\{\mu_T^\eta\}) \leq \frac{1}{b}$.*

(iv) *If the boundary condition η is free then $\kappa = \frac{e^{2\beta}-1}{e^{2\beta}+q-1}$.*

Part (i) of Theorem 5.29 follows from parts (i) and (ii) of Theorem 5.30 and the fact that for any boundary condition $\kappa \leq \gamma$. Part (ii) of Theorem 5.29 follows from parts (ii) and (iii) of Theorem 5.30 and the fact that $\gamma < 1$ (as is apparent from part (i) of the same theorem). Finally, part (iii) of Theorem 5.29 follows from parts (i) and (iv) of Theorem 5.30. In fact, in light of the bound on γ from part (i), the range of β in part (iii) of Theorem 5.29 can be improved slightly by letting β_1 be the solution to the equation $\frac{e^{2\beta_1}-1}{e^{2\beta_1}+q-1}(\frac{e^{2\beta_1}-1}{e^{2\beta_1}+1} - \delta) = \frac{1}{b}$, where $\delta = \delta(b, q, \beta)$ is as in part (i) of Theorem 5.30. We note that this modified definition of β_1 is only marginally larger than the original definition of β_1 in Theorem 5.29, and we mention it only in order to show that we can go further than the original threshold, a fact that is interesting due to its implication for the reconstruction problem [Mos02] as mentioned above.

Proof of Theorem 5.30: Much as we did for the previously discussed models, the first step we take in order to bound κ and γ is expressing the influence of a boundary spin as a function of the distribution of its neighboring interior spin, when the boundary spin is free. Generalizing Proposition 5.20 from the Ising model to the Potts model gives:

Proposition 5.31 *For any subset $A \subseteq T$, any boundary configuration τ, any pair of spins (s_1, s_2), any site $y \in \partial A$ and any neighbor $z \in A$ of y, we have*

$$\|\mu_A^{\tau^{y,s_1}} - \mu_A^{\tau^{y,s_2}}\|_z = K(p(s_1), p(s_2)),$$

where $p(s) = \mu_A^{\tau^{y,*}}(\sigma_z = s)$ and the function K is defined by

$$K(p_1, p_2) = \max\left\{ \frac{e^{2\beta}p_1}{(e^{2\beta}-1)p_1 + 1} - \frac{p_1}{(e^{2\beta}-1)p_2 + 1}\,,\, \frac{e^{2\beta}p_2}{(e^{2\beta}-1)p_2 + 1} - \frac{p_2}{(e^{2\beta}-1)p_1 + 1} \right\}, \tag{5.25}$$

and we notice that $K(p_1, p_2)$ is the first term in the maximum if and only if $p_1 \geq p_2$.

Proof: Let $p^s(s') = \mu_A^{\tau^{y,s}}(\sigma_z = s')$. Then $p^s(s) = \frac{e^{2\beta}p(s)}{e^{2\beta}p(s)+1-p(s)} = \frac{e^{2\beta}p(s)}{(e^{2\beta}-1)p(s)+1}$, and for $s' \neq s$, $p^s(s') = \frac{p(s')}{e^{2\beta}p(s)+1-p(s)} = \frac{p(s')}{(e^{2\beta}-1)p(s)+1}$, by definition of the Potts model. Now, the proposition follows by noticing that $\|\mu_A^{\tau^{y,s_1}} - \mu_A^{\tau^{y,s_2}}\|_z = \max\{p^{s_1}(s_1) - p^{s_2}(s_1), p^{s_2}(s_2) - p^{s_1}(s_2)\}$. The reason for this equality is that if $p_1 \geq p_2$ then $p^{s_1}(s) \leq p^{s_2}(s)$ for all $s \neq s_1$ (and $p^{s_1}(s_1) \geq p^{s_2}(s_1)$), as a simple calculation verifies. □

We can now easily dispense with parts (ii) and (iv) of Theorem 5.30. For part (iv) we simply observe that, for η the free boundary condition, the distribution $p(\cdot)$ at the root of the tree is uniform, i.e., $p(s) = \frac{1}{q}$ for all s, and therefore $\kappa = K(\frac{1}{q}, \frac{1}{q}) = \frac{e^{2\beta}-1}{e^{2\beta}+q-1}$, as required. For part (ii), observe that when the Gibbs measure is unique the distribution of the spin at the root of the tree converges as the depth increases to the uniform distribution, uniformly in the boundary condition. Thus, if A includes the full subtree of depth d rooted at z then, for every s, $p(s)$ converges with d to $\frac{1}{q}$, i.e., $\widehat{\gamma} \leq K(\frac{1}{q}+\varepsilon', \frac{1}{q}+\varepsilon') = K(\frac{1}{q}, \frac{1}{q})+\varepsilon \leq \frac{e^{2\beta}-1}{e^{2\beta}+q-1}+\varepsilon$ for some ε' and ε that go to zero with d, as required. (Notice that the Gibbs measure is unique in the regime $\beta < \beta_0$, for which $\frac{e^{2\beta}-1}{e^{2\beta}+q-1} < \frac{1}{b}$.)

We go on to prove part (i). Here we have to consider an arbitrary boundary configuration, and we cannot assume the distribution $p(\cdot)$ is close to uniform because the bound should also apply when the Gibbs distribution is not unique. Thus, the approach we take is simply calculating the maximum of $K(p_1, p_2)$ over all possible distributions. As a first step, let $K_{\max}(a) = \max_{p_1,p_2}\{K(p_1, p_2) : p_1 + p_2 = a\}$ denote the maximum restricted to distributions in which the sum of the two probabilities is a. We now observe that $K_{\max}(a)$ is strictly increasing in a. This is a consequence of the following two facts. First, $K(p, p)$ is strictly increasing in p. Second, for $p_1 > p_2$, $K(p_1, p_3) > K(p_1, p_2)$ for every $p_3 \in (p_2, p_1]$. We thus conclude that $\max_{p_1,p_2} K(p_1, p_2) = K_{\max}(1) = \max_p \left\{\frac{e^{2\beta}p}{(e^{2\beta}-1)p+1} - \frac{p}{(e^{2\beta}-1)(1-p)+1}\right\}$. It is now easy to verify that the expression in the maximization on the r.h.s. coincides with K_β defined for the Ising model in Proposition 5.20, under the change of variables p to $R = \frac{1-p}{p}$, and that this expression is maximized for $p = \frac{1}{2}$. We therefore conclude that $\gamma \leq \max_{p_1,p_2} K(p_1, p_2) = K(\frac{1}{2}, \frac{1}{2}) = \frac{e^{2\beta}-1}{e^{2\beta}+1}$. Now, recall that in part (i) of Theorem 5.30 we claimed a sharper bound on γ for $q > 2$. Specifically, we claimed that we can improve on the last bound by δ, where $\delta \equiv \delta(b, q, \beta)$ increases in q and decreases in b and β. This follows from the observation that, for all subsets A and boundary conditions τ, p_1 and p_2 in Proposition 5.31 are such that $p_1 + p_2 < 1$, and hence, since $K_{\max}(a)$ is strictly increasing in a, $K(p_1, p_2) < K_{\max}(1) = \frac{e^{2\beta}-1}{e^{2\beta}+1}$. The reason that $p_1+p_2 < 1$ is that, for any spin s, $p(s) >$

0 for all subsets A and boundary conditions τ. In particular, $p(s) > \frac{1}{e^{2\beta(b+1)}+q-1}$ for every s, i.e., $p(s_1) + p(s_2) \leq 1 - \frac{q-2}{e^{2\beta(b+1)}+q-1}$. We thus conclude that $\gamma \leq K_{\max}(1 - \frac{q-2}{e^{2\beta(b+1)}+q-1})$. Since $K_{\max}(a)$ increases with a, we observe that $\delta = K_{\max}(1) - K_{\max}(1 - \frac{q-2}{e^{2\beta(b+1)}+q-1})$ indeed increases with q and decreases with b and β.

We now go on to prove part (iii) of Theorem 5.30 (the last remaining part). Here the boundary is constant. W.l.o.g. we assume it is all 1. In order to bound κ we need to consider the distribution $p(\cdot)$ of the spin at the root of maximal subtrees with boundary condition all 1 at the bottom. Notice that by symmetry, $p(s)$ is uniform in $s \neq 1$. Therefore, this distribution is completely specified by $p(1)$ since for $s \neq 1$, $p(s) = \frac{1-p(1)}{q-1}$. As we shall see below, the fact that the distribution at the root is one-parameter allows for an analysis that is similar to that carried out for two-spin systems in Section 5.3.3.

To start, notice that by Proposition 5.31, $\|\mu^1_{T_z} - \mu^s_{T_z}\| = K(p(1), \frac{1-p_z(1)}{q-1})$, where we recall that $\mu^s_{T_z} = \mu^{\eta^{y,s}}_{T_z}$, and where $p_z(s) = \mu^{\eta^{y,*}}_{T_z}(\sigma_z = s)$. Similarly, for s_1, s_2 both different from 1, $\|\mu^{s_1}_T - \mu^{s_2}_T\| = K(\frac{1-p_z(1)}{q-1}, \frac{1-p_z(1)}{q-1})$. We now observe that, since the system is ferromagnetic, $p_z(1) \geq \frac{1}{q}$ for all z. An explicit calculation reveals that for every $p \geq \frac{1}{q}$, $K(p, \frac{1-p}{q-1}) \geq K(\frac{1-p}{q-1}, \frac{1-p}{q-1})$. Thus in order to bound κ, it is enough to bound $K(p_z(1), \frac{1-p_z(1)}{q-1})$ for every z. It is now convenient to consider the ratio $R = \frac{1-p(1)}{p(1)}$ and define

$$K(R) \equiv K\left(p(1), \frac{1-p(1)}{q-1}\right) = \frac{e^{2\beta}}{e^{2\beta}+R} - \frac{1}{(\frac{e^{2\beta}+q-2}{q-1})R+1}\,. \tag{5.26}$$

Observe that $\kappa = \sup_T \sup_z K(R_z)$. Thus, we need to show that this supremum is at most $\frac{1}{b}$. We now use the fact that the distribution at the root is one-parameter once again, this time to recursively calculate R_z, as was done for two-spin systems. In particular, we notice that $R_z = (q-1)\prod_{w \prec z} F(R_w)$, where $F(a) = \frac{(\frac{e^{2\beta}+q-2}{q-1})a+1}{a+e^{2\beta}}$. As in the two-spin case, we let $J(a) = (q-1)[F(a)]^b$ and observe that $R_z = J^{(\ell)}(0)$, where ℓ is the height of z. Notice that the functions $K(a)$ and $J(a)$ as defined here correspond exactly to the same two functions, defined for the two spin-system given by $(\lambda, \lambda_{(-)}, \lambda_{(+)}) = (q-1, \frac{e^{2\beta}+q-2}{q-1}, e^{2\beta})$. In other words, if we translate the spin 1 of the Potts model to $(+)$ and any non-1 spin to $(-)$, the Potts model with all-1 boundary corresponds exactly to the above two-spin system with all-$(+)$ boundary. To get some intuition for this translation, notice that $\lambda = q-1$ stems from the fact that a $(-)$-spin represents $q-1$ different spins (uniformly weighted) of the Potts model. For the same reason, $\lambda_{(-)} = \frac{e^{2\beta}+q-2}{q-1}$ stems from the fact that, given a non-1 spin, the interaction with another non-1 spin is composed of a fraction $1/(q-1)$ times $e^{2\beta}$ (when

the neighboring site is of the same Potts spin), and a fraction $(q-2)/(q-1)$ times 1 (for the interaction with the other $q-2$ spins).

Given this correspondence, it is now clear that κ for the all-1 boundary in the Potts model (the supremum over $K(R_z)$ given above) is exactly κ for the all-$(+)$ boundary in the two-spin system. Thus, we can conclude the proof of Theorem 5.30 part (iii) by using Theorem 5.25 (ii) for the two-spin system, once we show that non-uniqueness of the Gibbs measure for the Potts model implies non-uniqueness of the Gibbs measure in the two-spin system.

Now, if the Gibbs measure is not unique for Potts then, conditioned on the all-1 boundary configuration, the probability that the spin at the root is 1 is $> 1/q$ for arbitrarily large trees (since the model is ferromagnetic). This implies that in the two-spin system the probability of $(+)$ is $> 1/q$, which immediately implies non-uniqueness in this system because $a_* = q-1$ is always a fixed point of the function J given above. (In fact, it is not too difficult to see that the reverse implication holds as well, i.e., that the Gibbs measure is unique for the Potts model with parameters (q, β) if and only if it is unique for the corresponding two-spin system.)

We conclude that, for the Potts model in the regime of non-uniqueness of the Gibbs measure and for a constant boundary configuration, $\kappa \leq \frac{1}{b}$. This completes the proof of Theorem 5.30 part (iii). □

5.3.6 General q-spin systems

We end this section with an open question. In Section 5.3.3 we saw that for any two-spin system, if the Gibbs measure is unique then $\widehat{\gamma} \leq \frac{1}{b} + \varepsilon$ (and thus c_{sob} is bounded uniformly in the boundary condition). In Sections 5.3.4 and 5.3.5, we saw that the same holds for two examples of systems with $q > 2$ spins, where we used some symmetry of the systems to show this. A natural question is whether the above is true for any spin system with a finite spin space $\mathcal{S}$. We conjecture this to be true:

Conjecture 5.32 *For any q-spin system on the b-ary tree (where b and q are arbitrary), if the system admits a unique Gibbs measure then $\widehat{\gamma} \leq \frac{1}{b} + \varepsilon$, where ε goes to zero with the implicit parameter d in the definition of $\widehat{\gamma}$; in particular, if the Glauber dynamics for the system is connected then c_{sob} is bounded uniformly in the boundary condition.*

A possible direction for solving the above conjecture is analyzing the multivariate recursion for the distribution at the root of the tree. It is reasonable to expect that the total variation distance in $\widehat{\gamma}$ can be expressed in terms of the derivative of the function defining the recursive step, if the derivative is taken along an appropriate direction. Furthermore, we believe this derivative to "behave well" when the Gibbs measure is unique because the latter expresses a form of "attractiveness" of the fixed-point, where here a point is a probability vector.

Chapter 6

Boundary-specific mixing on the square integer lattice

In this chapter we investigate a few possible directions for establishing rapid (i.e., polynomial in the volume n) mixing time of the Glauber dynamics for specific boundary conditions by appealing to certain spatial mixing properties of the Gibbs distribution conditioned on these boundary configurations; in Chapter 5 we did this for systems on trees (where $O(m \log n)$ mixing time for specific boundary conditions was established), and here our we focus on the square lattice $\mathbb{Z}^2$. Our main motivation is the classical Ising model at low temperature with the all-$(+)$ boundary condition. As we already discussed in Section 5.1.1, the mixing time in this setting is conjectured to be bounded by a fixed polynomial at all temperatures, but no rigorous proof of this is known. The discussion in this chapter sheds more light on the problem and suggests a few possible directions towards solving it.

Our discussion of mixing in time in this chapter is focused on c_{gap}. Recall that by Theorem 2.9 (i), the mixing time is bounded by $O(mn \cdot c_{\text{gap}}^{-1})$. Also, it is standard (and easy to see) that c_{gap}^{-1} is bounded by the mixing time (normalized by m). Hence, mixing time bounded by a fixed polynomial in n is equivalent to c_{gap} bounded below by a fixed inverse polynomial in n (though of course these polynomials will be different).

The rest of the chapter is organized as follows. In Section 6.1 we set the context for our discussion by giving a detailed review of the state-of-the-art of the Ising model at low temperatures with $(+)$-boundary, for mixing in both time and space. In Section 6.2 we describe the general framework that the theory in this chapter is based on. This framework

is essentially the theory of variance decomposition discussed in Appendix A.2 and its relationship to c_{gap}. In Section 6.3 we present and prove a correspondence between mixing in time ("large" c_{gap}) and mixing in space that is boundary specific and holds in any bipartite graph. The spatial mixing condition we present in this section is new, and expresses the property that correlations between the "odd" and "even" subsets are "not too strong". Although we are not sure of the extent to which this condition is useful (i.e., how easy it is to verify it), it serves the purpose of establishing a two-way correspondence between mixing in time and space that is specific to the boundary condition. In Section 6.4 we specialize the discussion to the square lattice, where the geometry (essentially its planarity) allows for a correspondence between c_{gap} and a more relaxed spatial mixing condition. The relaxed condition expresses the property that correlations between two subsets at distance $O(\log n)$ from each other are not too strong. This latter condition seems more promising, since an appropriate decay of correlations implies it. However, this condition is still stronger than the decay of correlations that is currently known to take place on the square lattice at low temperatures with $(+)$-boundary condition. We elaborate on this point in the final Section 6.5, where we put the theory presented in Sections 6.3 and 6.4 in the context of the known and conjectured properties of the Ising model with $(+)$-boundary at low temperatures, and where we discuss future directions suggested by this theory for resolving the above conjecture.

6.1 The Ising model at low temperatures with $(+)$-boundary condition

In this section we describe some known facts concerning the Ising model on $\mathbb{Z}^2$ at low temperatures and conditioned on the all-$(+)$ boundary condition. This setting was already discussed in Section 5.1.1, but here we go into further detail.

Recall that the Ising model on $\mathbb{Z}^2$ has a critical inverse temperature β_c such that the Gibbs measure is unique if and only if $\beta \leq \beta_c$. Here we consider the case $\beta > \beta_c$, where multiple Gibbs measures exist and, in particular, the all-$(+)$ and all-$(-)$ boundary conditions yield significantly different Gibbs distributions. As already mentioned in Section 5.1.1, the mixing time of the Glauber dynamics in a square of volume n with the free-boundary condition is $\exp[\Theta(\sqrt{n})]$. (This follows from the fact that $c_{\mathrm{gap}} = \exp[-\Theta(\sqrt{n})]$ [CGMS96].)

We recall that the small c_{gap} in the free-boundary case is due to the bi-modal shape of the Gibbs distribution, where w.h.p. there is a majority of either $(+)$-spins or $(-)$-spins, and the probability of balanced configurations is exponentially small (in fact, $\exp[-\Theta(\sqrt{n})]$). Since conditioning on the all-$(+)$ boundary eliminates the $(-)$-phase and hence this bottleneck, it has been conjectured (e.g., in [Mar98, FH87]) that with the all-$(+)$ boundary, the mixing time of the Glauber dynamics should remain bounded by a fixed polynomial at all temperatures. This captures the intuition that the above bottleneck is the only obstacle for fast mixing of the Glauber dynamics. However, formalizing this intuition has so far proved to be very elusive.

We go on to describe some of the currently known facts regarding the Ising model with $(+)$-boundary at low temperatures. We start with the Gibbs distribution. It is known ([Dob96], and also [GHM01]) that the Gibbs distribution conditioned on the all-$(+)$ boundary exhibits exponential decay of correlations as in Definition 2.5. Specifically, for all $\beta > \beta_c$, there exist constants C and $\alpha > 0$ depending only on β such that, for any two functions f_A and g_B that depend only on A and B respectively, for every Ψ that includes $A \cup B$, and for η the all-$(+)$ configuration,

$$\text{Cov}^{\eta}_{\Psi}(f_A, g_B) \leq C \min\{|A|, |B|\} \, |f_{A,\max} - f_{A,\min}||g_{B,\max} - g_{B,\min}| \, \exp[-\alpha \cdot \text{dist}(A, B)], \tag{6.1}$$

where $f'_{\max}$ and $f'_{\min}$ stand for the maximum and minimum values of a function f' respectively. This exponential decay of correlations is another reason for the mixing time conjecture, since as we have seen throughout this thesis, mixing in time and mixing in space are intimately related.

We conclude the section with known results regarding the mixing time in the above setting. It is known that the mixing time for the all-$(+)$ boundary is faster than for the free boundary, but the improved bound is still exponential in a power of n. Specifically, it is known [HW99] that c_{gap} for the all-$(+)$ boundary is bounded below by $\exp[-n^{1/4}(\log n)^2]$. In terms of lower bounds, on integer lattices of any dimension d the mixing time for $\beta > \beta_c(d)$ and any boundary condition is clearly $\Omega(mn^{1/d})$. The reason is that if the mixing time for some boundary condition were $o(mn^{1/d})$, then it would follow from arguments similar to those given in Chapter 4 that the Gibbs measure is unique, a contradiction. (The uniqueness of the Gibbs measure follows from Lemma 4.4 in a similar manner to that used in the proof of Theorem 4.6, i.e., if the mixing time is $o(mn^{1/d})$, then there is not enough time for the

spin at the origin to be affected by the boundary condition.) Upper bounds on $c_{\rm gap}$ and $c_{\rm sob}$ (and hence lower bounds on the mixing time) for the all-$(+)$ boundary were recently given in [BM02]. Specifically, it was shown that in two dimensions $c_{\rm gap} = \widetilde{O}(n^{-1/2})$ (where $\widetilde{O}(\cdot)$ hides factors polylogarithmic in n), and that in all dimensions $d \geq 2$, $c_{\rm sob} = \widetilde{O}(n^{-2/d})$. These bounds are conjectured to be tight, and in particular, it is conjectured that $c_{\rm gap} = \Theta(n^{-1/2})$ in two dimensions for all $\beta > \beta_c$, and $c_{\rm gap} = \Theta(1)$ in dimensions three and higher at all temperatures.

6.2 The general framework

Throughout this chapter the only dynamics we discuss is Glauber (i.e., updates are made to single sites), and we do this by analyzing the spectral gap of the dynamics. Thus, as in Chapter 5, we write $c_{\rm gap}(\mu_\Psi^\eta)$ for $c_{\rm gap}(P)$, where P is the Glauber dynamics for sampling from μ_Ψ^η. Our approach is based on the theory of variance decomposition given in Appendix A.2. Consider two subsets A, B such that $A \cup B = \Psi$. In Appendix A.2 we define the quantity

$$\mathcal{V}_{\{A,B\}} \;=\; \inf_f \frac{\mu_\Psi^\eta[\mathrm{Var}_A(f)] + \mu_\Psi^\eta[\mathrm{Var}_B(f)]}{\mathrm{Var}_\Psi^\eta(f)},$$

where the infimum is over non-constant functions f. As discussed in Appendix A.2, $\mathcal{V}_{\{A,B\}}$ (which measures the well-decomposition of variance into the sum of conditional variance in A and B) is related to lack of correlations in the Gibbs distribution μ_Ψ^η, between the configurations on $A \setminus B$ and $B \setminus A$ respectively. The measure for correlations between two non-intersecting subsets A, B given in the same appendix is

$$\mathcal{C}_{\{A,B\}} \;=\; \sup_{f_A, g_B} \frac{\mathrm{Cov}_\Psi^\eta(f,g)^2}{\mathrm{Var}_\Psi^\eta(f_A)\mathrm{Var}_\Psi^\eta(g_B)},$$

where the supremum is over non-constant functions f_A and g_B that depend only on A and B respectively. Notice that $\mathcal{C}_{\{A,B\}}$ is at most 1, and that the smaller $\mathcal{C}_{\{A,B\}}$ is, the less correlated are the configurations on A and B. (See Appendix A.2.) Theorem A.4 in the same appendix establishes a correspondence between $\mathcal{V}_{\{A,B\}}$ and $\mathcal{C}_{\{A\setminus B, B\setminus A\}}$ that we will use throughout this chapter, i.e, the content of our spatial mixing conditions will be that $\mathcal{C}_{\{A\setminus B, B\setminus A\}}$ is "not too close" to 1 for a certain pair of subsets A, B that cover Ψ, and this corresponds to $\mathcal{V}_{\{A,B\}}$ being "not too close" to 0.

Bounding $\mathcal{V}_{\{A,B\}}$ is relevant for $c_{\rm gap}$ of the Glauber dynamics because of the following standard bound (e.g., [Mar98]), the derivation of which we repeat here for completeness. For any two subsets A, B such that $A \cup B = \Psi$,

$$\begin{aligned}
c_{\rm gap}(\mu_\Psi^\eta) &= \inf_f \frac{\sum_{x\in\Psi} \mu_\Psi^\eta[\mathrm{Var}_{\{x\}}(f)]}{\mathrm{Var}_\Psi^\eta(f)} \\
&= \inf_f \frac{\mu_\Psi^\eta[\mathrm{Var}_A(f)] + \mu_\Psi^\eta[\mathrm{Var}_B(f)]}{\mathrm{Var}_\Psi^\eta(f)} \cdot \frac{\sum_{x\in\Psi} \mu_\Psi^\eta[\mathrm{Var}_{\{x\}}(f)]}{\mu_\Psi^\eta[\mathrm{Var}_A(f)] + \mu_\Psi^\eta[\mathrm{Var}_B(f)]} \\
&\geq \mathcal{V}_{\{A,B\}} \cdot \tfrac{1}{2} \inf_f \frac{\sum_{x\in A} \mu_\Psi^\eta[\mathrm{Var}_{\{x\}}(f)] + \sum_{x\in B} \mu_\Psi^\eta[\mathrm{Var}_{\{x\}}(f)]}{\mu_\Psi^\eta[\mathrm{Var}_A(f)] + \mu_\Psi^\eta[\mathrm{Var}_B(f)]} \\
&= \tfrac{1}{2}\mathcal{V}_{\{A,B\}} \cdot \inf_f \frac{\mu_\Psi^\eta[\mu_A[\sum_{x\in A} \mathrm{Var}_{\{x\}}(f)]] + \mu_\Psi^\eta[\mu_B[\sum_{x\in B} \mathrm{Var}_{\{x\}}(f)]]}{\mu_\Psi^\eta[\mathrm{Var}_A(f)] + \mu_\Psi^\eta[\mathrm{Var}_B(f)]} \\
&\geq \tfrac{1}{2}\mathcal{V}_{\{A,B\}} \cdot \inf_f \min\left\{ \min_\tau \frac{\sum_{x\in A} \mu_A^\tau[\mathrm{Var}_{\{x\}}(f)]}{\mathrm{Var}_A^\tau(f)]}, \min_\tau \frac{\sum_{x\in B} \mu_B^\tau[\mathrm{Var}_{\{x\}}(f)]}{\mathrm{Var}_B^\tau(f)]} \right\} \\
&\geq \tfrac{1}{2}\mathcal{V}_{\{A,B\}} \cdot \min\left\{ \min_\tau c_{\rm gap}(\mu_A^\tau), \min_\tau c_{\rm gap}(\mu_B^\tau) \right\}, \qquad (6.2)
\end{aligned}$$

where in the minimum over τ, the latter is restricted to be a configuration in the support of μ_Ψ^η. Notice that the factor $\frac{1}{2}$ is only necessary if A and B overlap, and can be omitted if A and B are a partition of Ψ.

Following (6.2), in order to bound $c_{\rm gap}(\mu_\Psi^\eta)$ it is enough to bound $\mathcal{V}_{\{A,B\}}$ and $\min\{\min_\tau c_{\rm gap}(\mu_A^\tau), \min_\tau c_{\rm gap}(\mu_B^\tau)\}$. A bound on $\mathcal{V}$ can be obtained as described above using a spatial mixing assumption; specifically, since by Theorem A.4(ii) $\mathcal{V}_{\{A,B\}} \geq (1 - \mathcal{C}_{\{B\setminus A, A\setminus B\}})^2$, then by plugging this bound into (6.2) we get

$$c_{\rm gap}(\mu_\Psi^\eta) \;\geq\; \tfrac{1}{2}(1 - \mathcal{C}_{\{B\setminus A, B\setminus A\}})^2 \min\left\{ \min_\tau c_{\rm gap}(\mu_A^\tau), \min_\tau c_{\rm gap}(\mu_B^\tau) \right\}. \qquad (6.3)$$

Thus, if we can bound $\min\{\min_\tau c_{\rm gap}(\mu_A^\tau)\,,\, \min_\tau c_{\rm gap}(\mu_B^\tau)\}$, a bound on $c_{\rm gap}$ in terms of $\mathcal{C}$ will follow. In scenarios where the spatial mixing holds uniformly in the boundary condition, this minimum can be bounded by recursively applying (6.2), i.e., $c_{\rm gap}(\mu_A^\tau)$ is recursively bounded by considering two subsets A', B' such that $A' \cup B' = A$, and such that $\mathcal{C}_{\{A'\setminus B', B'\setminus A'\}}$ is small, where $\mathcal{C}$ is w.r.t. the Gibbs distribution μ_A^τ. This approach was taken in, e.g., [BCC02, Ces01, DPP02], to obtain bounded $c_{\rm gap}$ and $c_{\rm sob}$ uniformly in the boundary condition under a strong spatial mixing assumption for systems on the integer lattice. Notice, however, that recursive arguments of this type require that the spatial mixing assumption hold for arbitrary boundary conditions τ, while we are interested in deriving a bound on $c_{\rm gap}$ assuming spatial mixing *only* for the specific boundary configuration η. Thus,

rather than continuing recursively, we will work with subsets A, B for which there are "good enough" direct bounds on $c_{\rm gap}(\mu_A^\tau)$ and $c_{\rm gap}(\mu_B^\tau)$, uniformly in τ.

We conclude this section with the observation that bounds in the reverse direction to (6.2) and (6.3) exist; namely, that $\mathcal{V}_{A,B}$ and $\mathcal{C}_{B\setminus A, A\setminus B}$ are bounded in terms of $c_{\rm gap}(\mu_\Psi^\eta)$. This follows from the fact that $\mathrm{Var}_A(f) \geq \mathrm{Var}_{\{x\}}(f)$ for every function f and $x \in A$ (see, e.g., (A.1) in the appendix). Thus,

$$\begin{aligned} \mathcal{V}_{\{A,B\}} &= \inf_f \frac{\mu_\Psi^\eta[\mathrm{Var}_A(f)] + \mu_\Psi^\eta[\mathrm{Var}_B(f)]}{\mathrm{Var}(f)} \\ &\geq \inf_f \frac{\sum_{x\in A} \mu_\Psi^\eta[\mathrm{Var}_{\{x\}}(f)] + \sum_{x\in B} \mu_\Psi^\eta[\mathrm{Var}_{\{x\}}(f)]}{\max\{|A|,|B|\}\,\mathrm{Var}(f)} \\ &\geq n^{-1} c_{\rm gap}(\mu_\Psi^\eta)\,, \end{aligned} \tag{6.4}$$

where $n = |\Psi|$. Notice that by Theorem A.4(i), (6.4) yields

$$\mathcal{C}_{\{A\setminus B, B\setminus A\}} \;\leq\; 1 - n^{-1} c_{\rm gap}(\mu_\Psi^\eta)\,. \tag{6.5}$$

We note that if $c_{\rm gap}$ is bounded and if $\mathrm{dist}(A\setminus B\,,\, B\setminus A)$ is large w.r.t. $\min\{|A|,|B|\}$, then a much better bound exists. Specifically, in [KMP01] it was shown that

$$\mathcal{C}_{\{A\setminus B, B\setminus A\}} \;\leq\; C \min\{|A|,|B|\} \exp[-\vartheta c_{\rm gap}(\mu_\Psi^\eta) \mathrm{dist}(A\setminus B\,,\, B\setminus A)] \tag{6.6}$$

for some constants C and $\vartheta > 0$ that depend only on the potentials of the system and the maximum degree of G. However, the bound in (6.6) is non-trivial only if $c_{\rm gap}(\mu_\Psi^\eta) \cdot \mathrm{dist}(A \setminus B\,,\, B\setminus A) \geq \Omega(\log(C \min\{|A|,|B|\}))$. For example, the bound is trivial if $c_{\rm gap}(\mu_\Psi^\eta) = o(n^{-1})$. The advantage of the bound in (6.5) is that it is always non-trivial.

6.3 A boundary-specific space-time correspondence for bipartite graphs

In this section we consider spin systems on bipartite graphs $G = G_1 \cup G_2$, and prove a two-way correspondence between $c_{\rm gap}$ of the Glauber dynamics and correlations between odd and even sites in the Gibbs distribution. Our arguments are based on the trivial observation that, conditioned on the configuration on the odd (respectively, even) sites, the Gibbs distribution on the even (respectively, odd) sites is the product of its marginals over single sites, where by "even" and "odd" we mean sites in G_1 and G_2 respectively. Our rather simple result reads as follows:

Theorem 6.1 *Consider an arbitrary Ψ and boundary condition η, and let Ψ_1, Ψ_2 be the partition of Ψ into even and odd sites respectively. Then,*

(i) $c_{\mathrm{gap}}(\mu_\Psi^\eta) \geq (1 - \mathcal{C}_{\{\Psi_1,\Psi_2\}})^2$;

(ii) $\mathcal{C}_{\{\Psi_1,\Psi_2\}} \leq 1 - n^{-1} c_{\mathrm{gap}}(\mu_\Psi^\eta)$.

Notice that this is a correspondence between a spatial mixing notion ($\mathcal{C}_{\{\Psi_1,\Psi_2\}}$) that depends only on the distribution μ_Ψ^η, and a temporal mixing notion ($c_{\mathrm{gap}}(\mu_\Psi^\eta)$) that depends only on the same distribution. In particular, an immediate consequence of Theorem 6.1 is that $c_{\mathrm{gap}}(\mu_\Psi^\eta)$ is bounded below by an inverse polynomial in n if and only if $\mathcal{C}_{\Psi_1,\Psi_2}$ is bounded away from 1 by an inverse polynomial in n.

Proof: Part (ii) is just a special case of (6.5). Part (i) will follow from (6.3) once we show that $c_{\mathrm{gap}}(\mu_{\Psi_i}^\tau) = 1$ for all τ and $i = 1, 2$. (Notice that we omitted the factor $\frac{1}{2}$ from (6.3) because Ψ_1, Ψ_2 is a partition of Ψ.) Now, observe that for every τ and $i = 1, 2$, the distribution $\mu_{\Psi_i}^\tau$ is the product of its marginals over single sites (since the sites in Ψ_i are isolated once the spins in the other part are fixed), and hence $c_{\mathrm{gap}}(\mu_{\Psi_i}^\tau) = 1$. The fact that $c_{\mathrm{gap}} = 1$ for product distributions is standard (e.g., [Sal97]) and was already mentioned in Section 5.2.1, where we used the fact that for a product distribution μ_Λ^τ, variance decomposes perfectly into the sum of conditional variances in any partition of Λ. □

Remark: The fact that $c_{\mathrm{gap}} = 1$ for a product distribution μ_Λ^τ can also be proven by applying (6.3) recursively, and by noticing that $\mathcal{C}_{\{A,B\}} = 0$ for any partition A, B of Λ when μ_Λ^τ is a product distribution (and again recalling that the factor $\frac{1}{2}$ in (6.3) is not needed if the two subsets do not overlap).

6.4 A boundary-specific correspondence for the square lattice

In this section we continue the theme started in the previous section and establish a correspondence between $\mathcal{C}_{\{A\setminus B, B\setminus B\}}$ and c_{gap} for a certain choice of the subsets A, B. Recall that in Theorem 6.1, the subsets A, B are the odd and even subsets respectively, and the distance between $(A \setminus B) = A$ and $(B \setminus A) = B$ is 1. We would like to establish a correspondence of this type where the distance between $A \setminus B$ and $B \setminus A$ is larger, in the hope that this will

make it easier to establish a small $\mathcal{C}_{\{A\setminus B, B\setminus A\}}$. In this section we indeed consider pairs of subsets with a larger distance between them, but our discussion is restricted to the square lattice $\mathbb{Z}^2$, whose geometry (in particular, planarity) allows for an appropriate choice of the subsets A, B.

Fix an arbitrary $\Psi \subset \mathbb{Z}^2$. For every positive integer ℓ, let

$$\begin{aligned} \mathcal{A}_\ell &= \left\{x = (x_1, 2a\ell) \in \mathbb{Z}^2 \ : \ x_1, a \in \mathbb{Z}, x \in \Psi\right\}; \\ \mathcal{B}_\ell &= \left\{x = (x_1, (2a+1)\ell) \in \mathbb{Z}^2 \ : \ x_1, a \in \mathbb{Z}, x \in \Psi\right\} \end{aligned}$$

denote the intersections with Ψ of the sets of lines at even and odd multiples of ℓ respectively. Set $A_\ell = \Psi \setminus \mathcal{A}_\ell$ and $B_\ell = \Psi \setminus \mathcal{A}_\ell$. Clearly, for every ℓ, $(A_\ell \cup B_\ell) = \Psi$. Notice also that $A_\ell \setminus B_\ell = \Psi \setminus B_\ell = \mathcal{B}_\ell$. Similarly, $B_\ell \setminus A_\ell = \mathcal{A}_\ell$. We establish the following correspondence between $c_{\rm gap}$ and $\mathcal{C}_{\{\mathcal{A}_\ell, \mathcal{B}_\ell\}}$ for systems on $\mathbb{Z}^2$:

Theorem 6.2 *There exists a constant $\vartheta > 0$ depending only on the potentials of the system such that, for every region $\Psi \subset \mathbb{Z}^2$, any boundary configuration η and all ℓ:*

(i) $c_{\rm gap}(\mu_\Psi^\eta) \geq (1 - \mathcal{C}_{\mathcal{A}_\ell, \mathcal{B}_\ell})^2 n^{-1} e^{-\vartheta \ell}$;

(ii) $\mathcal{C}_{\mathcal{A}_\ell, \mathcal{B}_\ell} \leq 1 - n^{-1} c_{\rm gap}(\mu_\Psi^\eta)$,

where $\mathcal{C}_{\mathcal{A}_\ell, \mathcal{B}_\ell}$ is defined w.r.t. μ_Ψ^η and $n = |\Psi|$.

Notice that $\mathrm{dist}(\mathcal{A}_\ell, \mathcal{B}_\ell) = \ell$, i.e., $\mathcal{C}_{\mathcal{A}_\ell, \mathcal{B}_\ell}$ is a measure of correlations between subsets at distance ℓ from each other. Thus, the condition in the above theorem is flexible enough to consider situations where correlations decay with distance, but a certain minimum distance is required to see this effect. In particular, it follows from the theorem that if $\mathcal{C}_{\mathcal{A}_\ell, \mathcal{B}_\ell}$ is bounded away from 1 by an inverse polynomial (in n) for some $\ell = O(\log n)$, then $c_{\rm gap}$ is bounded below by an inverse polynomial.

Proof: Part (ii) is again just a reiteration of (6.5) and the only reason for stating it is for the correspondence to read in both directions. For part (i) we once again use (6.3), and thus it is enough to show that $\max\left\{\max_\tau c_{\rm gap}(\mu_{A_\ell}^\tau)\,,\, \max_\tau c_{\rm gap}(\mu_{A_\ell}^\tau)\right\} \geq n^{-1} e^{-\vartheta\ell}$, for some constant $\vartheta > 0$ that depends only on the potentials of the system. We will show only that $c_{\rm gap}(\mu_{A_\ell}^\tau) \geq n^{-1} e^{-\vartheta\ell}$ for every τ since the argument for B_ℓ is identical. The main observation we use is that A_ℓ is a disjoint union of clusters $A_{\ell,i}$, where each $A_{\ell,i}$ is a region of width $2\ell - 1$, delimited by the lines at even multiples of ℓ. (Each $A_{\ell,i}$ is isolated

from the rest of the clusters — see figure 6.1.) It is standard that $c_{\rm gap}$ of a finite union of isolated clusters is the least $c_{\rm gap}$ of a single cluster, i.e., $c_{\rm gap}(\mu^\tau_{A_\ell}) = \min_i c_{\rm gap}(\mu^\tau_{A_{\ell,i}})$. (This follows from the fact that the Gibbs distribution $\mu^\tau_{A_\ell}$ is the product of its marginals over the clusters $A_{\ell,i}$.) We thus go on to bound $c_{\rm gap}(\mu^\tau_{A_{\ell,i}})$. Now, it is well known (see, e.g., [Mar98]) that there exists a constant $\vartheta' > 0$ depending only on the potentials of the system such that, for every subset Λ of width b and any τ, $c_{\rm gap}(\mu^\tau_\Lambda) \geq |\Lambda|^{-1} e^{-\vartheta'(b+1)}$. This follows from a flow along canonical-paths argument [Sin92] and the fact that the "cut-width" of Λ is at most $b+1$, i.e., there exists an enumeration $x_1, \ldots, x_{|\Lambda|}$ of the sites in Λ such that, for every j, the number of edges connecting the two subsets $\{x_1, \ldots, x_j\}$ and $\{x_{j+1}, \ldots, x_{|\Lambda|}\}$ is at most $b+1$. (See [Mar98] for details.) We conclude that, for every τ and i, $c_{\rm gap}(\mu^\tau_{A_{\ell,i}}) \geq |A_{\ell,i}|^{-1} e^{-\vartheta' 2\ell} \geq n^{-1} e^{-2\vartheta'\ell}$, and therefore $c_{\rm gap}(\mu^\tau_{A_\ell}) \geq n^{-1} e^{-2\vartheta'\ell}$ for every τ, completing the proof of part (i). □

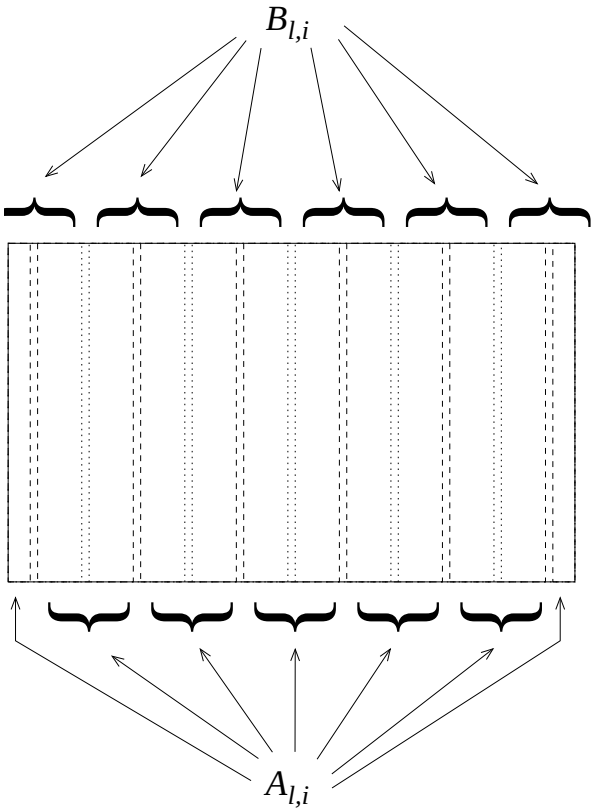

Figure 6.1: The regions A_ℓ and B_ℓ are composed of the clusters $A_{\ell,i}$ and $B_{\ell,i}$ respectively. The clusters $A_{\ell,i}$ (respectively, $B_{\ell,i}$) are separated by the lines at even (respectively, odd) multiples of ℓ.

6.5 The theory in the context of the Ising model on $\mathbb{Z}^2$

In this section we compare the decay of correlations that is known to take place in the Ising model on $\mathbb{Z}^2$ at low temperatures with all-$(+)$boundary condition, to the conditions on correlations that, by Theorems 6.1 and 6.2, would imply $c_{\rm gap}$ inverse polynomial in n.

As mentioned in Section 6.1, it is known that the Ising model in this setting exhibits the exponential decay of correlations specified in (6.1). Notice that this does not give any information on the correlations between the odd and even subsets $\mathcal{C}_{\{\Psi_1,\Psi_2\}}$ in Theorem 6.1 because the distance between the two subsets is 1 and their size is order of n. We will discuss future directions for research involving $\mathcal{C}_{\{\Psi_1,\Psi_2\}}$ at the of the section; for now let us proceed with relating the decay of correlations expressed in (6.1) to $\mathcal{C}_{\{\mathcal{A}_\ell,\mathcal{B}_\ell\}}$ in Theorem 6.2. As discussed immediately after that theorem, in order to lower bound $c_{\rm gap}$ by an inverse polynomial in n, it is enough to show that $\mathcal{C}_{\{\mathcal{A}_\ell,\mathcal{B}_\ell\}}$ is bounded away from 1 by an inverse polynomial in n for some $\ell = O(\log n)$. (This is also a necessary condition.) Now, for two subsets A, B separated by a distance $\Omega(\log(\min\{|A|,|B|\})) = \Omega(\log n)$, (6.1) (the decay that is known to take place) gives non-trivial bounds for functions whose L_∞ norm is on the order of their L_2 norm. Indeed, if we could replace the L_∞ normalization on the r.h.s. of (6.1) with an L_2 normalization (i.e., replace $|f_{A,\max} - f_{A,\min}||g_{B,\max} - g_{B,\min}|$ with $\sqrt{\mathrm{Var}^\eta_\Psi(f)\mathrm{Var}^\eta_\Psi(g)}$) then it would follow that $\mathcal{C}_{\mathcal{A}_\ell,\mathcal{B}_\ell}$ is bounded away from 1 for some $\ell = O(\log n)$, and hence that $c_{\rm gap}$ is bounded below by an inverse polynomial in n. We also mention that (6.1) implies

$$\mathrm{Cov}^\eta_\Psi(f_A, g_B) \le C\sqrt{\mathrm{Var}^\eta_\Psi(f_A)\mathrm{Var}^\eta_\Psi(g_B)}\,\exp[\alpha_1 \min\{|A|,|B|\} - \alpha_2 \cdot \mathrm{dist}(A,B)], \qquad (6.7)$$

once we notice that the minimum non-zero probability of a configuration on A is at least inverse exponential in $|A|$. However, (6.7) gives non-trivial bounds only for A, B such that $\mathrm{dist}(A,B) = \Omega(\min\{|A|,|B|\})$, i.e., it gives non-trivial bounds on $\mathcal{C}_{\{\mathcal{A}_\ell,\mathcal{B}_\ell\}}$ only for $\ell = \Omega(n)$, which is much larger than the ℓ we need to work with.

We summarize the above discussion with the observation that $c_{\rm gap}$ bounded below by an inverse polynomial in n would follow from the following form of exponential decay of correlations:

$$\mathrm{Cov}^\eta_\Psi(f_A, g_B) \le \mathrm{poly}(|A|,|B|)\sqrt{\mathrm{Var}^\eta_\Psi(f_A)\mathrm{Var}^\eta_\Psi(g_B)}\,\exp[-\alpha \cdot \mathrm{dist}(A,B)], \qquad (6.8)$$

for some constant $\alpha > 0$ that depends only on β and for any two functions f_A and f_B that

depend only on the configurations on A and B respectively. Thus, one suggested direction for future research is establishing (6.8) for large β (low temperatures) and η all-$(+)$.

We also mention that, from the above discussion, it is apparent that the constants C_f in Definition 2.5 of exponential decay of correlations play a crucial role. As already mentioned, for the Ising model on $\mathbb{Z}^2$ at low temperatures with η all-$(+)$ this decay holds with an appropriate choice of constants as expressed in (6.1). We now show that this kind of exponential decay of correlations does not hold in the same setting for other choices of the constants C_f. Specifically, we show that the decay does not hold for $C_f = C\sqrt{\mathrm{Var}^{\eta}_{\Psi}(f)}$, where C is an arbitrary constant uniform in f. (Notice that this C_f does not depend on the size of the subset that f depends on.) Suppose for the sake of contradiction that, for any two non-intersecting subsets A, B and any pair of functions f_A, g_B that depend only A and B respectively,

$$\mathrm{Cov}^{\eta}_{\Psi}(f_A, g_B) \leq C\sqrt{\mathrm{Var}^{\eta}_{\Psi}(f_A)\mathrm{Var}^{\eta}_{\Psi}(g_B)}\, \exp[-\alpha \cdot \mathrm{dist}(A, B)], \tag{6.9}$$

for every $\Psi \supseteq A \cup B$, η all-$(+)$ and for some constants C and $\alpha > 0$ that depend only on β. It then follows that $\mathcal{C}_{\{\mathcal{A}_\ell, \mathcal{B}_\ell\}} < 1$ uniformly in Ψ for some constant ℓ (independent of $n = |\Psi|$). Now, for ℓ bounded, it is not too difficult to obtain a strengthening of Theorem 6.2(i), where the factor n^{-1} is omitted. Hence, $\mathcal{C}_{\{\mathcal{A}_\ell, \mathcal{B}_\ell\}} < 1$ for a bounded ℓ implies that c_{gap} is bounded for η, a contradiction. (Recall from Section 6.1 that it is known [BM02] that, on $\mathbb{Z}^2$, c_{gap} is not bounded for η all-$(+)$ at low temperatures.) The above strengthening of Theorem 6.2(i) follows from the fact that $c_{\mathrm{gap}}(\mu^{\tau}_{\mathcal{A}_\ell})$ is bounded independently of n if ℓ is bounded. This is because the clusters $\mathcal{A}_{\ell,i}$ in the proof of this theorem are then of bounded width, i.e., essentially one-dimensional, and since c_{gap} is always bounded for one-dimensional subsets.

We conclude the section with another suggestion for future research. The first suggestion we made was to establish (6.8) in the setting of the Ising model on $\mathbb{Z}^2$ with all-$(+)$boundary at low temperatures. Notice that (6.8) is not necessary for c_{gap} to be inverse polynomial in n. In fact, some researchers believe (6.8) is false because c_{gap} is not bounded in the above setting. (Currently, of course, there is no formal proof either way.) However, as is apparent from Theorems 6.1 and 6.2, there are other conditions on $\mathcal{C}$ that are not only sufficient for c_{gap} to be bounded below by an inverse polynomial, but are also necessary. Therefore, establishing these conditions on $\mathcal{C}$ is a natural direction for future research. More specifically, notice that $c_{\mathrm{gap}} = \Omega(\mathrm{poly}^{-1}(n))$ is equivalent to $\mathcal{C}_{\{\Psi_1, \Psi_2\}}$ bounded away from 1 by an inverse polynomial in n (Theorem 6.1), and to $\mathcal{C}_{\{\mathcal{A}_\ell, \mathcal{B}_\ell\}}$ similarly bounded for some

$\ell = O(\log n)$. This suggests a need for new tools for establishing a much more delicate form of "weak" correlations, where "weak" actually means "not too strong", and in particular, for establishing $\mathcal{C}_{\{A,B\}} \leq 1 - \text{poly}^{-1}(|A|, |B|)$ even if the distance between A and B is small. The current tools are much coarser, and only capable of establishing that $\mathcal{C}_{\{A,B\}}$ goes to zero with $\text{dist}(A, B)$, i.e., for large distances they give much stronger bounds, but no information for small distances. To put our suggestion in a broader context, we mention that it is not too difficult to see that $\mathcal{C}_{\{A,B\}} \leq 1 - \exp[-C \min\{|\partial A|, |\partial B|\}]$ for any pair of non-intersecting subsets A, B, where C is a constant that depends only on the potentials of the system and the maximum degree of the underlying graph G. This follows from the fact that $|g_{\{A,B\}}|_\infty \leq 1 - \exp[-C \min\{|\partial A|, |\partial B|\}]$, where $|g_{\{A,B\}}|_\infty$ is defined in Appendix A.3. We refer to [Mar98] for a derivation of a similar bound.

Bibliography

[ABCF*00] C. ANÉ, S. BLACHÈRE, D. CHAFAÏ, P. FOUGÈRES, I. GENTIL, F. MALRIEU, C. ROBERTO and G. SCHEFFER, "Sur les inégalités de Sobolev logarithmiques," *Société Mathématique de France*, 2000.

[AH87] M. AIZENMAN and R. HOLLEY, "Rapid convergence to equilibrium of stochastic Ising models in the Dobrushin Shlosman regime," in: H. Kesten, *Percolation Theory and Ergodic Theory of Infinite Particle Systems*, IMS Volumes in Math. and Appl. **8**, pp. 1–11, Springer-Verlag, New York (1987).

[Ald83] D. ALDOUS, "Random walks on finite groups and rapidly mixing Markov chains," *Séminaire de Probabilites XVII*, Lecture Notes in Mathematics **986**, pp. 243–297, Springer, Berlin, 1983.

[AP03] D. ACHLIOPTAS and Y. PERES, "The threshold for random k-SAT is $2^k \ln 2 - O(k)$," *Proceedings of the 35h ACM Symposium on Theory of Computing,* 2003, pp. 223–231.

[Bax82] R.J. BAXTER, *Exactly solved models in statistical mechanics*, Academic Press, London, 1982.

[BCC02] L. BERTINI, N. CANCRINI and F. CESI, "The spectral gap for a Glauber-type dynamics in a continuous gas," *Ann. Inst. H. Poincaré Probab. Statist.* **38** (2002), pp. 91–108.

[BD97] R. BUBLEY and M.E. DYER, "Path coupling: a technique for proving rapid mixing in Markov chains," *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science*, 1997, pp. 223–231.

[Ber93] J. van den Berg, A uniqueness condition for Gibbs measures, with applications to the 2-dimensional Ising antiferromagnet, *Comm. Math. Phys.*, **152**, 161-166 (1993).

[BM02] T. BODINEAU and F. MARTINELLI, "Some new results on the kinetic Ising model in a pure phase," *Journal of Statistical Physics* **109** (1), 2002.

[BMZ02] A. BRAUNSTEIN, M. MÉZARD, and R. ZECCHINA, "Survey propagation: an algorithm for satisfiability," Preprint, 2002.

[BRSSZ01] P. BLEHER, J. RUIZ, R.H. SCHONMANN, S. SHLOSMAN and V. ZAGREBNOV, "Rigidity of the critical phases on a Cayley tree," *Moscow Mathematical Journal* **1** (2001), pp. 345–363.

[BRZ95] P. BLEHER, J. RUIZ and V. ZAGREBNOV, "On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice," *Journal of Statistical Physics* **79** (1995), pp. 473–482.

[BW02] G. BRIGHTWELL and P. WINKLER, "Random colorings of a Cayley tree," *Contemporary combinatorics*, Bolyai Society Mathematical Studies **10**, János Bolyai Math. Soc., Budapest, 2002, pp. 247–276.

[BW03] G.R. BRIGHTWELL and P. WINKLER, "A second threshold for the hard-core model on a Bethe lattice," INI Report NI03002-CMP. To appear in *Rand. Struc. & Alg.*

[Ces01] F. CESI, "Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields," *Probability Theory and Related Fields* **120** (2001), pp. 569–584.

[CGMS96] F. CESI, G. GUADAGNI, F. MARTINELLI and R.H. SCHONMANN, "On the 2D stochastic Ising model in the phase coexistence region close to the critical point," *Journ. Stat. Phys.* **85** (1996), pp. 55-102.

[DF01] M. DYER and A. FRIEZE, "Randomly colouring graphs with lower bounds on girth and maximum degree," *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, 2001, pp. 579–587.

[DFPV04] M. DYER, A. FRIEZE, T.P. HAYES and E. VIGODA, "Randomly colouring constant degree graphs," Preprint, 2004.

[DG00] M. DYER and C. GREENHILL, "On Markov chains for independent sets," *J. Algorithms* **35** (2000), pp. 17–49.

[Dob70] R.L. DOBRUSHIN, "Prescribing a system of random variables by the help of conditional distributions," *Theory Prob. and its Appl.* **15** (1970), pp. 469–497.

[Dob96] R. DOBRUSHIN, " Estimates of semi-invariants for the Ising model at low temperatures," *Amer. Math. Soc. Transl.* **177** (1996), pp. 59–81.

[DPP02] P. DAI PRA, A. M. PAGANONI and G. POSTA, "Entropy inequalities for unbounded spin systems," to appear in *Annals of Probability.*

[DS85a] R.L. DOBRUSHIN and S.B. SHLOSMAN, "Constructive criterion for the uniqueness of a Gibbs field," in: J. FRITZ, A. JAFFE, D. SZASZ, *Statistical mechanics and dynamical systems*, Birkhauser, Boston (1985), pp. 347–370.

[DS85b] R.L. DOBRUSHIN and S.B. SHLOSMAN, "Completely Analytical Gibbs Fields," in: J. Fritz, A. Jaffe, D. Szasz, *Statistical Mechanics and Dynamical Systems*, pp. 371–403, Birkhauser, Boston (1985).

[DSVW02] M. DYER, A. SINCLAIR, E. VIGODA and D. WEITZ, "Mixing in time and space for lattice spin systems: A combinatorial view," *Proc. 6th International workshop on Randomization and Approximation Techniques (RANDOM)*, 2002, pp. 149–163. Full version to appear in a special issue of *Random Structures and Algorithms*.

[EKPS00] W. EVANS, C. KENYON, Y. PERES and L.J. SCHULMAN, "Broadcasting on trees and the Ising model," *Annals of Applied Probability* **10** (2000), pp. 410–433.

[FB99] E. FRIEDGUT, "Sharp thresholds of graph properties, and the k-SAT problem," with an appendix by J. BOURGAIN, *Journal of the American Mathematical Society* **12** (1999), pp. 1017–1054.

[FH87] D. FISHER and D. HUSE, "Dynamics of droplet fluctuations in pure and random Ising systems," *Physics Review B* **35** (13), 1987.

[FK99] A. FRIEZE and R. KANNAN, "Log-Sobolev inequalities and sampling from log-concave distributions," *Annals of Applied Probability* **9** (1999), pp. 14–26.

[FM01] J.A. FILL and M. MACHIDA, "Stochastic monotonicity and realizable monotonicity," *Annals of Probability* **29** (2001), pp. 938–978.

[Geo88] H.-O. GEORGII, *Gibbs measures and phase transitions*, de Gruyter Studies in Mathematics **9**, Walter de Gruyter & Co., Berlin, 1988.

[GHM01] H.-O. GEORGII, O.HÄGGSTRÖM and C. MAES, "The random geometry of equilibrium phases," *Phase Transitions and Critical Phenomena*, Volume 18 (C. Domb and J.L. Lebowitz, eds), Academic Press, London (2001), pp. 1–142.

[Hag96] O. HÄGGSTRÖM, "The random-cluster model on a homogeneous tree," *Probability Theory and Related Fields* **104** (1996), pp. 231–253.

[Hay03] T.P. HAYES, "Randomly coloring graphs with girth five," *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, 2003, pp. 269–278.

[Hol85] R. HOLLEY, "Possible rates of convergence in finite range, attractive spin systems," in: R. Durret, *Particle systems, Random media, and Large deviations*, AMS Volumes on Contemp. Math. **41** (1985), pp.215–234.

[HV03] T.P. HAYES and E. VIGODA, "A non-Markovian coupling for randomly sampling colorings," *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003, pp 618–627.

[HW99] Y. HIGUCHI and J. WANG, "Spectral gap of Ising model for Dobrushin's boundary condition in two dimension," Preprint, 1999.

[Iof96a] D. IOFFE, "A note on the extremality of the disordered state for the Ising model on the Bethe lattice," *Letters in Mathematical Physics* **37** (1996), pp. 137–143.

[Iof96b] D. IOFFE, "Extremality of the disordered state for the Ising model on general trees," *Progress in Probability* **40** (1996), pp. 3–14.

[Isi25] E. ISING, "Beitrag zur Theorie des Ferromagnetismus," *Zeitschrift für Physik* **31** (1925), pp. 253–258.

[Jon02] J. JONASSON "Uniqueness of uniform random colorings of regular trees," *Statistics & Probability Letters* **57** (2002), pp. 243–248.

[JS02] M. JERRUM and J.-B. SON, "Spectral gap and log-Sobolev constant for balanced matroids," *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science*, 2002, pp. 721–729.

[JS99] J. JONASSON and J.E. STEIF, "Amenability and phase transition in the Ising model," *Journal of Theoretical Probability* **12** (1999), pp. 549–559.

[JSTV02] M. JERRUM, J.-B. SON, P. TETALI and E. VIGODA, "Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains," Preprint, 2003.

[Kel85] F.P. KELLY, "Stochastic models of computer communication systems," *Journal of the Royal Statistical Society* B **47** (1985), pp. 379–395.

[KMP01] N. BERGER, C. KENYON, E. MOSSEL and Y. PERES, "Glauber dynamics on trees and hyperbolic graphs," Preprint, 2003. Preliminary version: C. KENYON, E. MOSSEL and Y. PERES, "Glauber dynamics on trees and hyperbolic graphs," *Proc. 42nd IEEE Symposium on Foundations of Computer Science* (2001), pp. 568–578.

[KRS89] A.B. KIRILLOV, D.C. RADULESCU and D.F. STYER, "Vassertein distances in two-state systems," *J. Stat. Phys.* **56** (1989), pp. 931–937.

[LV99] M. LUBY and E. VIGODA, "Fast convergence of the Glauber dynamics for sampling independent sets," *Random Structures & Algorithms* **15** (1999), pp. 229–241.

[Lyo89] R. LYONS, "The Ising model and percolation on trees and tree-like graphs," *Comm. Math. Phys.* **125** (1989), pp. 337–352.

[Lyo00] R. LYONS, "Phase transitions on non amenable graphs," *J. Math. Phys.* **41** (2000), pp. 1099–1127.

[Mar98] F. MARTINELLI, "Lectures on Glauber dynamics for discrete spin models," *Lectures on Probability Theory and Statistics (Saint-Flour, 1997)*, Lecture notes in Mathematics **1717**, pp. 93–191, Springer, Berlin, 1998.

[Mar03] J.B. MARTIN, "Reconstruction thresholds on regular trees," *http://arxiv.org/abs/math.PR/0305400*, to appear in DMTCS volume *Proc. Discrete Random Walks 2003*.

[MO94a] F. MARTINELLI and E. OLIVIERI, "Approach to equilibrium of Glauber dynamics in the one phase region I: The attractive case," *Communications in Mathematical Physics* **161** (1994), pp. 447–486.

[MO94b] F. MARTINELLI and E. OLIVIERI, "Approach to equilibrium of Glauber dynamics in the one phase region II: The general case," *Communications in Mathematical Physics* **161** (1994), pp. 487–514.

[Mol02] M. MOLLOY, "The Glauber dynamics on colorings of a graph with high girth and maximum degree," *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, 2002, pp. 91–98.

[MOS94] F. MARTINELLI, E. OLIVIERI and R. SCHONMANN, "For 2-D lattice spin systems weak mixing implies strong mixing," *Comm. Math. Phys.* **165** (1994), pp. 33–47.

[Mos02] E. MOSSEL, "Survey: Information flow on trees," Preprint, October 2002, to appear in DIMACS volume *Graphs, Morphisms and Statistical Physics*.

[MP01] E. MOSSEL and Y. PERES, "Information flow on trees," *Annals of Applied Probability* **13** (2003), pp. 817–844.

[MSW03] F. MARTINELLI, A. SINCLAIR and D. WEITZ, "Glauber dynamics on trees: Boundary conditions and mixing time," to appear in *Communications in Mathematical Physics*, 2004. Extended abstract appeared as: "The Ising Model on trees: Boundary conditions and mixing times," *Proceedings of the 44th IEEE Symposium on Foundations of Computer Science*, 2003, pp. 628–639.

[MSW04] F. MARTINELLI, A. SINCLAIR and D. WEITZ, "Fast mixing for independent sets, colorings and other models on trees," submitted, 2004. Extended abstract ap-

peared in: *Proceedings of the 15th ACM-SIAM Symposium on Discrete Algorithms*, 2004, pp. 449–458.

[Pot] R.B. POTTS, "Some generalized order-disorder transformations," *Proc. Cambridge Philos. Soc.* **48**, pp. 106–109.

[Pre74] C.J. PRESTON, *Gibbs states on countable sets*, Cambridge Tracts in Mathematics **68**, Cambridge University Press, London, 1974.

[PW02] Y. PERES and P. WINKLER, personal communication.

[Ric73] D. RICHARDSON, "Random growth in a tesselation," *Proc. Cambr. Phil. Soc.* **74** (1973), pp. 515–528.

[Sal97] L. SALOFF-COSTE, "Lectures on finite Markov chains," *Lectures on probability theory and statistics (Saint-Flour, 1996)*, Lecture notes in Mathematics **1665**, pp. 301–413, Springer, Berlin, 1997.

[Sim93] B. SIMON, *The statistical mechanics of lattice gases, Vol. I*, Princeton Series in Physics, Princeton University Press, Princeton, NJ, 1993.

[Sin92] A. SINCLAIR, "Improved bounds for mixing rates of Markov chains and multicommodity flow," *Combinatorics, Probability and Computing* **1** (1992), pp. 351–370.

[SJ89] A.J. SINCLAIR and M.R. JERRUM, "Approximate counting, uniform generation and rapidly mixing Markov chains," *Information and Computation* **82** (1989), pp. 93–133.

[Spi75] F. SPITZER, "Markov random fields on an infinite tree," *Annals of Probability* **3** (1975), pp. 387–398.

[SS97] J. SALAS and A.D. SOKAL, "Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem," *J. Stat. Phys.* **86** (1997), pp. 551–579.

[ST98] R.H. SCHONMANN and N.I. TANAKA, "Lack of monotonicity in ferromagnetic Ising model phase diagrams," *Annals of Applied Probability* **8** (1998), pp. 234–245.

[SZ92] D.W. STROOCK and B. ZEGARLINSKI, "The logarithmic Sobolev inequality for discrete spin systems on a lattice," *Communications in Mathematical Physics* **149** (1992), pp. 175–194.

[SZ95] D.W. STROOCK and B. ZEGARLINSKI, "On the ergodic properties of Glauber dynamics," *J. Statist. Phys.* **81** (1995), pp. 1007–1019.

[Vig00] E. VIGODA, "Improved bounds for sampling colorings," *Journal of Mathematical Physics* **41** (2000), pp. 1555–1569.

[Vig01] E. VIGODA, "A note on the Glauber dynamics for sampling independent sets," Electronic Journal of Combinatorics, Volume 8(1), 2001.

[Wei03] D. WEITZ "Combinatorial criteria for uniqueness of the Gibbs measure," submitted, 2003.

[Wu82] F.Y. WU, "The Potts model," *Reviews of Modern Physics* **54** (1982), pp. 235–268.

[Zeg90] B. ZEGARLINSKI, "On log-Sobolev inequalities for infinite lattice systems," *Lett. Math. Phys.* **20** (1990), pp. 173-182.

Appendix A

Decomposing variance and entropy

In this appendix we give a detailed discussion of decompositions of variance and entropy respectively, and in particular, of intimate relationships between these decompositions and lack of correlations in the Gibbs distribution. Relationships of this kind were used in the text, and here we give the proofs and references to previous results of this kind. Some of the theorems we give here were not used in the text; the reason for presenting them is to exhibit this decomposition theory in full, both because it helps in understanding the parts that were needed in the text, and because it might be useful for future research.

The discussion here is based on general notions from functional analysis one, and in fact holds in more general settings than spin systems, i.e., the analysis is of variance and entropy w.r.t. general distributions that are "close" to being product distributions in an appropriate sense.

Throughout the appendix we have a fixed distribution $\nu = \mu_\Psi^\eta$ in mind. We also consider the distribution on subsets A and B of Ψ. Namely, we write $\nu_A \equiv \mu_A$ for the conditional distribution in A. We abbreviate $\mathrm{Var}_\Psi^\eta(f)$ and $\mathrm{Ent}_\Psi^\eta(f)$ to $\mathrm{Var}(f)$ and $\mathrm{Ent}(f)$ respectively. We retain the original notation $\mathrm{Var}_A^\tau(f)$ (respectively $\mathrm{Ent}_A^\tau(f)$) when referring to the variance (respectively entropy) in a subset A of Ψ conditional on τ outside A, and as usual, $\mathrm{Var}_A(f)$ (respectively $\mathrm{Ent}_A(f)$) is the function representing the conditional variance (respectively entropy) of f in A. Since we consider Ψ to be fixed, we write $\overline{A} \equiv \Psi \setminus A$ for the complement of A in Ψ.

The rest of this appendix is organized as follows. We start by describing (and proving) a few basic properties of variance and entropy that are commonly used in our analysis. We then move on to establish specific relationships between lack of correlations in the dis-

tribution ν and closeness of the approximation of variance and entropy by a decomposition into local conditional terms.

A.1 Basic properties of variance and entropy

In this section we list (and prove for completeness) a few basic properties that are common to variance and entropy, and are useful for the rest of the discussion in this appendix.

We start with a trivial decomposition of the variance into the local conditional variance in a subset A and the variance of the projection outside A.

Proposition A.1 *For every function $f : \Omega \to \mathbb{R}$,*

$$\mathrm{Var}(f) \;=\; \nu[\mathrm{Var}_A(f)] + \mathrm{Var}[\nu_A(f)]. \tag{A.1}$$

The same holds when replacing Var *with* Ent *and restricting to non-negative f.*

Proof: For variance:

$$\begin{aligned}
\nu[\mathrm{Var}_A(f)] + \mathrm{Var}[\nu_A(f)] &= \nu[\nu_A(f^2) - \nu_A(f)^2] + \nu[\nu_A(f)^2] - \nu[\nu_A(f)]^2 \\
&= \nu[\nu_A(f^2)] - \nu[\nu_A(f)]^2 \\
&= \nu(f^2) - \nu(f)^2 = \mathrm{Var}(f).
\end{aligned}$$

For entropy:

$$\begin{aligned}
\mathrm{Ent}(f) &= \nu\left[f \log\left(\frac{f}{\nu(f)}\right)\right] \\
&= \nu\left[f \log\left(\frac{f}{\nu_A(f)}\right)\right] + \nu\left[f \log\left(\frac{\nu_A(f)}{\nu(f)}\right)\right] \\
&= \nu\left[\mathrm{Ent}_A(f)\right] + \nu\left[\nu_A(f) \log\left(\frac{\nu_A(f)}{\nu(f)}\right)\right] \\
&= \nu[\mathrm{Ent}_A(f)] + \mathrm{Ent}[\nu_A(f)]. \qquad \square
\end{aligned}$$

Our second observation is that variance (respectively entropy) w.r.t. a fixed distribution is a convex functional.

Proposition A.2 *For every convex combination $\sum_{i=1}^k \lambda_i = 1$ and k functions $f_1, \ldots, f_k$,*

$$\mathrm{Var}\left(\sum_{i=1}^k \lambda_i f_i\right) \;\le\; \sum_{i=1}^k \lambda_i \mathrm{Var}(f_i). \tag{A.2}$$

The same holds when replacing Var *with* Ent *and restricting to non-negative functions f_i.*

Proof: We first prove for Var:

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{k}\lambda_i f_i\right) &= \nu\left[\left(\sum_{i=1}^{k}\lambda_i f_i - \nu\left(\sum_{i=1}^{k}\lambda_i f_i\right)\right)^2\right] \\
&= \nu\left[\left(\sum_{i=1}^{k}\lambda_i (f_i - \nu(f_i))\right)^2\right] \\
&\leq \nu\left[\sum_{i=1}^{k}\lambda_i (f_i - \nu(f_i))^2\right] \\
&= \sum_{i=1}^{k}\lambda_i\, \nu\big[(f_i - \nu(f_i))^2\big] \\
&= \sum_{i=1}^{k}\lambda_i \mathrm{Var}(f_i),
\end{aligned}$$

where the inequality is Cauchy-Schwartz. We go on to prove this for entropy. Let $\alpha_i = \lambda_i\nu(f_i)/[\sum_{i=1}^{k}\lambda_i\nu(f_i)]$, and notice that $\sum_{i=1}^{k}\alpha_i = 1$. Then,

$$\begin{aligned}
\mathrm{Ent}\left(\sum_{i=1}^{k}\lambda_i f_i\right) &= \nu\left[\sum_{i=1}^{k}\lambda_i f_i \log\left(\frac{\sum_{i=1}^{k}\lambda_i f_i}{\nu(\sum_{i=1}^{k}\lambda_i f_i)}\right)\right] \\
&= \left(\sum_{i=1}^{k}\lambda_i\nu(f_i)\right)\cdot\nu\left[\left(\sum_{i=1}^{k}\alpha_i\frac{f_i}{\nu(f_i)}\right)\log\left(\sum_{i=1}^{k}\alpha_i\frac{f_i}{\nu(f_i)}\right)\right] \\
&\leq \left(\sum_{i=1}^{k}\lambda_i\nu(f_i)\right)\cdot\nu\left[\sum_{i=1}^{k}\alpha_i\frac{f_i}{\nu(f_i)}\log\left(\frac{f_i}{\nu(f_i)}\right)\right] \\
&= \nu\left[\sum_{i=1}^{k}\lambda_i f_i\log\left(\frac{f_i}{\nu(f_i)}\right)\right] \\
&= \sum_{i=1}^{k}\lambda_i \mathrm{Ent}(f_i),
\end{aligned}$$

where the inequality uses the fact that $x\log x$ is a convex function. □

We are now going to use the above convexity property to prove inequality (5.5) from Section 5.2.1. Although we do not need this inequality for the rest of the discussion in this appendix, we place the proof here because it expresses a certain decomposition property, and is based on a product structure of the Gibbs distribution, i.e., the fact that the distribution in A depends only on the configuration in ∂A, and is independent of the rest of the configuration once the boundary condition is fixed. (This is the only place in this appendix where the Markovian structure of the Gibbs distribution is used.)

Proposition A.3 *For any two subsets $A, B \subseteq \Psi$ such that $(\partial A) \cap B = \emptyset$, and for any function f,*

$$\nu[\mathrm{Var}_A(\nu_B(f))] \leq \nu[\mathrm{Var}_A(\nu_{A\cap B}(f))]. \tag{A.3}$$

The same holds when replacing Var *by* Ent *and restricting to non-negative f.*

Proof: We only prove the Var formulation because the only property of Var that we use in the proof is its convexity, which holds for Ent as well. First, for any function g and boundary configuration τ we have

$$\mathrm{Var}_A^\tau[\nu_{B\setminus A}(g)] \leq \nu_{B\setminus A}^\tau[\mathrm{Var}_A(g)], \tag{A.4}$$

as we now explain. Recall that by assumption $(\partial A) \cap B = \emptyset$, so there are no edges connecting the two disjoint subsets A and $B \setminus A$. Therefore, the distribution $\nu_{B\setminus A}$ does not depend on the configuration in A, and ν_A does not depend on the configuration in $B \setminus A$. Thus, if we fix the configuration outside A to be τ, then $\nu_{B\setminus A}(g)$ (a function only of the spins in A) can be written as

$$\nu_{B\setminus A}(g) = \sum_\sigma \nu_{B\setminus A}^\tau(\sigma) g_\sigma,$$

where g_σ is g with the spins on $B \setminus A$ fixed to σ. Note that this is a convex combination of functions g_σ. Therefore, we may use the fact that variance w.r.t. a fixed measure is a convex functional, together with the fact that the measure ν_A does not depend on the configuration on $B \setminus A$, to deduce

$$\mathrm{Var}_A^\tau[\nu_{B\setminus A}(g)] \leq \sum_\sigma \nu_{B\setminus A}^\tau(\sigma)\mathrm{Var}_A^\tau(g_\sigma) = \nu_{B\setminus A}^\tau[\mathrm{Var}_A(g)],$$

thus verifying (A.4).

Finally, equation (A.3) follows from (A.4) with $g = \nu_{B\cap A}(f)$ by writing

$$\begin{aligned}
\nu[\mathrm{Var}_A(\nu_B(f))] &= \nu[\mathrm{Var}_A(\nu_{B\setminus A}\nu_{B\cap A}(f))] \\
&\leq \nu[\nu_{B\setminus A}[\mathrm{Var}_A(\nu_{B\cap A}(f))]] \\
&= \nu[\mathrm{Var}_A(\nu_{B\cap A}(f))]. \qquad \square
\end{aligned}$$

A.2 Decomposing variance

In this section we establish relationships between lack of correlations in the distribution ν and well-behavedness of the decomposition of variance into local conditional terms. We

measure the well-decomposition of variance into that in (proper) subsets A, B that cover Ψ by the quantity $\mathcal{V}_{\{A,B\}}$, defined as follows:

$$\mathcal{V}_{\{A,B\}} \;\equiv\; \mathcal{V}_{\{A,B\}}(\nu) \;=\; \inf_f \frac{\nu[\mathrm{Var}_A(f)] + \nu[\mathrm{Var}_B(f)]}{\mathrm{Var}(f)}, \tag{A.5}$$

where the infimum is over non-constant functions f. Notice that $\mathcal{V}_{\{A,B\}} \leq 1$ because $\mathrm{Var}_A(f) \leq \mathrm{Var}(f)$ by (A.1) and we can always consider functions that do not depend on B, i.e., for which $\nu[\mathrm{Var}_B(f)] = 0$. Thus, we regard $\mathcal{V}_{\{A,B\}} = 1$ as the case in which variance decomposes perfectly into conditional variance in A and B, while $\mathcal{V}_{\{A,B\}} = 0$ means that the decomposition can be arbitrarily bad. Notice also that $\mathcal{V}_{\{A,B\}}$ is exactly c_{gap} of the dynamics based on flipping the two blocks A, B.

We measure correlations between two non-intersecting subsets by normalized covariance. Specifically for subsets A, B such that $A \cap B = \emptyset$, let

$$\mathcal{C}_{\{A,B\}} \;\equiv\; \mathcal{C}_{\{A,B\}}(\nu) \;=\; \sup_{f,g} \frac{\mathrm{Cov}(f,g)^2}{\mathrm{Var}(f)\cdot\mathrm{Var}(g)}, \tag{A.6}$$

where the supremum is over non-constant functions f and g that depend only on the configurations on A and B respectively. Notice that since f does not depend on $\overline{A}$ then

$$\mathrm{Cov}(f,g)^2 \;=\; \mathrm{Cov}[f, \nu_{\overline{A}}(g)]^2 \;\leq\; \mathrm{Var}(f)\cdot\mathrm{Var}[\nu_{\overline{A}}(g)],$$

with equality if $f = \nu_{\overline{A}}(g)$. Thus,

$$\mathcal{C}_{\{A,B\}} \;=\; \sup_g \frac{\mathrm{Var}[\nu_{\overline{A}}(g)]}{\mathrm{Var}(g)} \;=\; \sup_f \frac{\mathrm{Var}[\nu_{\overline{B}}(f)]}{\mathrm{Var}(f)}\,, \tag{A.7}$$

where the supremums are over non-constant functions g and f that depend only on the configurations on B and A respectively. In particular, $\mathcal{C}_{\{A,B\}} \leq 1$. Indeed, $\mathcal{C}_{\{A,B\}} = 0$ means that the configurations on A and B are independent of each other, while $\mathcal{C}_{\{A,B\}} = 1$ means that they are completely dependent, i.e., there exist non-trivial events $\mathcal{A}$ and $\mathcal{B}$ that depend only on the configurations on A and B respectively, such that $\mathcal{A}$ occurs if and only if $\mathcal{B}$ occurs. Notice that $\mathcal{C}_{\{A,B\}}$ being exponentially small in $\mathrm{dist}(A,B)$ corresponds to exponential decay of correlations (as in Definition 2.5) with normalizing constants $C_f = \sqrt{\mathrm{Var}(f)}$. We also mention that, by (A.7), $\mathcal{C}_{\{A,B\}}$ can be viewed as the contraction in variance when projecting a function that depends only on the configuration B onto A (or vice versa), and in particular, the variance mixing condition $\mathrm{VM}(\ell,\varepsilon)$ (as in Definition 5.7) that was used in the tree setting is equivalent to $\mathcal{C}_{\{A,B\}} = \varepsilon$, where $A = \{x\}$ and $B = T_x \setminus B_{x,\ell}$ (with notation as in Section 5.2).

Notice that $\mathcal{V}_{\{A,B\}}$ is defined for two subsets A, B that cover Ψ (and may intersect), while $\mathcal{C}_{\{A,B\}}$ is defined for two non-intersecting subsets (that may not cover Ψ). As we shall now see, $\mathcal{V}_{\{A,B\}}$ being close to 1 corresponds to $\mathcal{C}_{\{A\setminus B, B\setminus A\}} = \mathcal{C}_{\{\overline{B},\overline{A}\}}$ being close to 0, i.e., well-decomposition of variance corresponds to lack of correlations.

Theorem A.4 *For any two subsets A, B that cover Ψ:*

(i) $\mathcal{C}_{\{\overline{B},\overline{A}\}} \leq 1 - \mathcal{V}_{\{A,B\}}$;

(ii) $\mathcal{V}_{\{A,B\}} \geq \left(1 - \sqrt{\mathcal{C}_{\{\overline{B},\overline{A}\}}}\right)^2$.

Bounds of the above type were already established in [BCC02], and the proof below is based on arguments given there. However, our analysis for part (ii) is sharper and improves on the corresponding claim in [BCC02], which gives $\mathcal{V}_{\{A,B\}} \geq \left(1 - \sqrt{\mathcal{C}_{\{\overline{B},\overline{A}\}}}\right)^2 - 2\mathcal{C}_{\{\overline{B},\overline{A}\}}$. The main importance of our improvement is that it gives non-trivial bounds for all possible values of $\mathcal{C}$, while the result in [BCC02] is non-trivial only for $\mathcal{C} < 3 - 2\sqrt{2} < 1$.

Proof of Theorem A.4: We start with the rather trivial part (i). Consider an arbitrary function f that depends only on the configuration on $\overline{B}$. By (A.7), it is enough to show that $\mathrm{Var}[\nu_{\Psi\setminus\overline{A}}(f)] = \mathrm{Var}[\nu_A(f)] \leq (1 - \mathcal{V}_{\{A,B\}})\mathrm{Var}(f)$. Now clearly, $\nu[\mathrm{Var}_B(f)] = 0$ because f does not depend on the configuration on B. Therefore, by definition of $\mathcal{V}_{\{A,B\}}$, $\nu[\mathrm{Var}_A(f)] \geq \mathcal{V}_{\{A,B\}}\mathrm{Var}(f)$. Since $\mathrm{Var}[\nu_A(f)] = \mathrm{Var}(f) - \nu[\mathrm{Var}_A(f)]$ (by (A.1)), we conclude that $\mathrm{Var}[\nu_A(f)] \leq (1-\mathcal{V}_{\{A,B\}})\mathrm{Var}(f)$, as required. This completes the proof of part (i).

We go on to prove the more interesting part (ii). We first notice that $\nu[\mathrm{Var}_B(f)] = \mathrm{Var}[f - \nu_B(f)]$. To see this, write

$$\mathrm{Var}[f - \nu_B(f)] \;=\; \mathrm{Cov}[f - \nu_B(f), f - \nu_B(f)] \;=$$

$$\mathrm{Cov}(f,f) - 2\mathrm{Cov}[f, \nu_B(f)] + \mathrm{Cov}[\nu_B(f), \nu_B(f)] \;=\; \mathrm{Var}(f) - \mathrm{Var}[\nu_B(f)] \;=\; \nu[\mathrm{Var}_B(f)].$$

Consider an arbitrary function f. We need to show that $\nu[\mathrm{Var}_A(f)] + \nu[\mathrm{Var}_B(f)] \geq \left(1 - \sqrt{\mathcal{C}_{\{\overline{B},\overline{A}\}}}\right)^2 \mathrm{Var}(f)$. Now w.l.o.g., assume $\mathrm{Var}[\nu_A(f)] \geq \mathrm{Var}[\nu_B(f)]$. We will show that under this assumption,

$$\nu[\mathrm{Var}_B(f)] \;\geq\; \left(1 - \sqrt{\mathcal{C}_{\{\overline{B},\overline{A}\}}}\right)^2 \mathrm{Var}[\nu_A(f)]. \tag{A.8}$$

This is enough because then

$$\begin{aligned}\nu[\mathrm{Var}_A(f)] + \nu[\mathrm{Var}_B(f)] &\geq \left(1 - \sqrt{\mathcal{C}_{\{\overline{B},\overline{A}\}}}\right)^2 (\nu[\mathrm{Var}_A(f)] + \mathrm{Var}[\nu_A(f)]) \\ &= \left(1 - \sqrt{\mathcal{C}_{\{\overline{B},\overline{A}\}}}\right)^2 \mathrm{Var}(f).\end{aligned}$$

We go on to establish (A.8):

$$\begin{aligned}\mathrm{Var}[\nu_A(f)] &= \mathrm{Cov}[f, \nu_A(f)] \\ &= \mathrm{Cov}[f - \nu_B(f), \nu_A(f)] + \mathrm{Cov}[\nu_B(f), \nu_A(f)] \\ &\leq \mathrm{Cov}[f - \nu_B(f), \nu_A(f)] + \sqrt{\mathcal{C}_{\{\overline{B},\overline{A}\}} \cdot \mathrm{Var}[\nu_B(f)] \cdot \mathrm{Var}[\nu_A(f)]} \\ &\leq \mathrm{Cov}[f - \nu_B(f), \nu_A(f)] + \sqrt{\mathcal{C}_{\{\overline{B},\overline{A}\}}} \cdot \mathrm{Var}[\nu_A(f)] \\ &\leq \sqrt{\nu[\mathrm{Var}_B(f)] \cdot \mathrm{Var}[\nu_A(f)]} + \sqrt{\mathcal{C}_{\{\overline{B},\overline{A}\}}} \cdot \mathrm{Var}[\nu_A(f)],\end{aligned}$$

where for the first inequality we used the definition of $\mathcal{C}_{\{\overline{B},\overline{A}\}}$ and the fact that $\nu_A(f)$ and $\nu_B(f)$ depend only on the configurations on $\overline{A}$ and $\overline{B}$ respectively, for the second inequality we used the assumption that $\mathrm{Var}[\nu_A(f)] \geq \mathrm{Var}[\nu_B(f)]$, and the last inequality is Cauchy-Schwartz. We conclude that $\left(1 - \sqrt{\mathcal{C}_{\{\overline{B},\overline{A}\}}}\right) \sqrt{\mathrm{Var}[\nu_A(f)]} \leq \sqrt{\nu[\mathrm{Var}_B(f)]}$, i.e., (A.8) holds. This concludes the proof of part (ii). □

Theorem A.4 tells us that for every function f, $\mathrm{Var}(f) \leq (1 - \sqrt{\mathcal{C}_{\{A,B\}}})^{-2} \times (\nu[\mathrm{Var}_A(f)] + \nu[\mathrm{Var}_A(f)])$. However, we can also get an approximation with a tighter factor in front of $\nu[\mathrm{Var}_A(f)]$ by letting the factor in front of $\nu[\mathrm{Var}_B(f)]$ be looser. This is useful for getting a decomposition of $\mathrm{Var}[\nu_A(f)] = \mathrm{Var}(f) - \nu[\mathrm{Var}_A(f)]$ into $c \times \nu[\mathrm{Var}_B(f)] + \varepsilon \times \nu[\mathrm{Var}_A(f)]$, where we do not need the sharpest possible value for c (the best possible is $c = 1$), but look for a better ε than that given by Theorem A.4 (the best possible is $\varepsilon = \mathcal{C}_{\{\overline{B},\overline{A}\}}/(1 - \mathcal{C}_{\{\overline{B},\overline{A}\}})$). An approximation of this kind was used in Section 5.2.1. We show:

Theorem A.5 *Consider two arbitrary subsets A, B that cover Ψ, and let $\varepsilon = \mathcal{C}_{\{\overline{B},\overline{A}\}}$. Then, for every function f,*

$$\mathrm{Var}[\nu_A(f)] \leq \frac{2(1-\varepsilon)}{1-2\varepsilon} \cdot \nu[\mathrm{Var}_B(f)] + \frac{2\varepsilon}{1-2\varepsilon} \cdot \nu[\mathrm{Var}_A(f)].$$

Proof: By convexity of variance we have $\mathrm{Var}(g_1+g_2) \leq 2[\mathrm{Var}(g_1)+\mathrm{Var}(g_2)]$ for any two functions g_1, g_2. We therefore write

$$\begin{aligned}
\mathrm{Var}[\nu_A(f)] &= \mathrm{Var}[\nu_A(f)-\nu_A(\nu_B(f))+\nu_A(\nu_B(f)] \\
&\leq 2\mathrm{Var}[\nu_A(f-\nu_B(f))]+2\mathrm{Var}[\nu_A(\nu_B(f))] \\
&\leq 2\mathrm{Var}[f-\nu_B(f)]+2\varepsilon\mathrm{Var}[\nu_B(f)] \\
&= 2\nu[\mathrm{Var}_B(f)]+2\varepsilon(\mathrm{Var}[\nu_A(f)]+\nu[\mathrm{Var}_A(f)]-\nu[\mathrm{Var}_B(f)]),
\end{aligned}$$

where we used the facts that $\mathrm{Var}[f-\nu_B(f)] = \nu[\mathrm{Var}_B(f)]$ and that $\mathrm{Var}[\nu_A(f)]+\nu[\mathrm{Var}_A(f)] = \mathrm{Var}[\nu_B(f)]+\nu[\mathrm{Var}_B(f)] = \mathrm{Var}(f)$. We therefore conclude that $\mathrm{Var}[\nu_A(f)] \leq \frac{2(1-\varepsilon)}{1-2\varepsilon}\cdot\nu[\mathrm{Var}_B(f)]+\frac{2\varepsilon}{1-2\varepsilon}\cdot\nu[\mathrm{Var}_A(f)]$, as required. □

Before concluding this section, we wish to shed more light on the quantity $\mathcal{C}_{\{A,B\}}$ by bounding it using specific functions. Recall that σ_A stands for the restriction of the configuration σ to the subset A. For σ_A in the support of ν, define the function

$$g_{\sigma_A}(\xi) = \begin{cases} \frac{1}{\nu(\sigma_A)} & \text{if } \xi_A = \sigma_A; \\ 0 & \text{otherwise,} \end{cases} \tag{A.9}$$

and for two non-intersecting subsets, define $g_{(A,B)}(\sigma,\xi) \equiv \nu^{\xi}_{B}(g_{\sigma_A})$, i.e., $g_{(A,B)}(\sigma,\xi) = \frac{\nu^{\xi}_{B}(\sigma_A)}{\nu(\sigma_A)}$ measures how much σ_A is biased by conditioning on ξ_B. By Bayes rule, $\frac{\nu^{\xi}_{B}(\sigma_A)}{\nu(\sigma_A)} = \frac{\nu^{\sigma}_{A}(\xi_B)}{\nu(\xi_B)}$, and therefore $g_{(A,B)}(\sigma,\xi) = g_{(B,A)}(\xi,\sigma)$. Notice that, if the distribution ν is the product of its marginals over A and B, then $g_{(A,B)} = 1$ uniformly in both configurations. More generally, the concentration of $g_{(A,B)}$ around it mean value 1 is a measure of lack of correlations between A and B. Indeed, $\mathcal{C}_{\{A,B\}}$ is bounded by the average (w.r.t the first configuration) of the variance (w.r.t. the second configuration) of $g_{(A,B)}$ (though the order can be reversed, by symmetry of $\mathcal{C}_{\{A,B\}}$ and $g_{(A,B)}$):

Theorem A.6 *For any two non-intersecting subsets A, B,*

$$\mathcal{C}_{\{A,B\}} \leq \sum_{\sigma}\nu(\sigma)\mathrm{Var}[g_{(A,B)}(\sigma,\cdot)].$$

We note that Theorem A.6 is not needed for the main text, but is rather standard and sets a benchmark for concentration properties of g, to which we can compare when relating such properties to the well-decomposition of entropy in the next section.

Proof of Theorem A.6: Consider an arbitrary function f that depends only on the configuration on B. By (A.7), it is enough to show that $\mathrm{Var}[\nu_{\overline{A}}(f)] \leq \varepsilon \mathrm{Var}(f)$, where $\varepsilon = \sum_\sigma \nu(\sigma)\mathrm{Var}[g_{(A,B)}(\sigma,\cdot)]$. The main observation we use is that, since f and $g_{(A,B)}(\sigma,\cdot) = g_{(B,A)}(\cdot,\sigma)$ depend only on B, then for every σ,

$$\begin{aligned}\nu^\sigma_{\overline{A}}(f) &= \sum_{\xi_B} \nu^\sigma_{\overline{A}}(\xi_B)\cdot f(\xi_B) \;=\; \sum_{\xi_B} \nu(\xi_B) g_{(B,A)}(\xi_B,\sigma)\cdot f(\xi_B) \\ &= \sum_{\xi_B} \nu(\xi_B) g_{(A,B)}(\sigma,\xi_B)\cdot f(\xi_B) \;=\; \nu[g_{(A,B)}(\sigma,\cdot)\, f(\cdot)].\end{aligned}$$

Using in addition the fact that $\nu[g_{(A,B)}(\sigma,\cdot)] = 1$ for every σ gives:

$$\begin{aligned}\mathrm{Var}(\nu_{\overline{A}}(f)) &= \nu[(\nu_{\overline{A}}(f) - \nu(f))^2] \\ &= \sum_\sigma \nu(\sigma)\cdot(\nu^\sigma_{\overline{A}}(f) - \nu(f))^2 \\ &= \sum_\sigma \nu(\sigma)\cdot \mathrm{Cov}[g_{(A,B)}(\sigma,\cdot)\,,\; f(\cdot)]^2 \qquad \text{(A.10)} \\ &\leq \mathrm{Var}(f)\sum_\sigma \nu(\sigma)\cdot\mathrm{Var}[g_{(A,B)}(\sigma,\cdot)]\,,\end{aligned}$$

where the inequality is Cauchy-Schwartz. □

A.3 Decomposing entropy

In this section we carry out an analysis analogous to the one in the previous section, by replacing variance with entropy. Namely, we relate the well-decomposition of entropy to certain properties of lack of correlations in the distribution ν.

The well-decomposition of entropy into that in (proper) subsets A, B that cover Ψ is measured by a quantity analogous to $\mathcal{V}_{A,B}$. Specifically, define

$$\mathcal{E}_{\{A,B\}} \;\equiv\; \mathcal{E}_{\{A,B\}}(\nu) \;=\; \inf_{f\geq 0} \frac{\nu[\mathrm{Ent}_A(f)] + \nu[\mathrm{Ent}_B(f)]}{\mathrm{Ent}(f)}, \qquad \text{(A.11)}$$

where the infimum is over non-constant and non-negative functions f. It is easy to see that $\mathcal{E}_{\{A,B\}} \leq 1$ for the same reason that $\mathcal{V}_{\{A,B\}} \leq 1$.

We consider two types of measures of correlations between two non-intersecting subsets. The first is an analog of $\mathcal{C}$ based on the characterization given in (A.7). Specifically, for two subsets A, B such that $A \cap B = \emptyset$, define

$$\mathcal{N}_{(A,B)} \;\equiv\; \mathcal{N}_{(A,B)}(\nu) \;=\; \sup_{f\geq 0} \frac{\mathrm{Ent}[\nu_{\overline{A}}(f)]}{\mathrm{Ent}(f)}, \qquad \text{(A.12)}$$

where the supremum is over non-negative non-constant functions f that depend only on the configuration on B. Thus, $\mathcal{N}_{(A,B)}$ measures the (weakest) contraction in entropy when projecting a function that depends only on the configuration in B onto A. Indeed, this is a measure of correlation between the configurations on A and B respectively. For example, if the two configurations are independent of each other then $\mathcal{N}_{(A,B)} = 0$, while if they are completely dependent then $\mathcal{N}_{(A,B)} = 1$. Notice that unlike $\mathcal{C}_{\{A,B\}}$, $\mathcal{N}_{(A,B)}$ is not necessarily symmetric in A and B.

The second type of measure of correlations that we consider are certain concentration properties of the function $g_{(A,B)}$ defined at the end of the previous section. (See the discussion after (A.9).) As a warm up, we bound $\mathcal{E}_{\{A,B\}}$ in terms of the uniform norm of $g_{(A,B)}$. Define

$$|g_{\{A,B\}}|_\infty = \inf\left\{\varepsilon \;:\; 1-\varepsilon \;\le\; g_{(A,B)}(\sigma,\xi) \;\le\; \frac{1}{1-\varepsilon} \text{ for every } \sigma,\xi \text{ in the support of } \nu\right\}, \tag{A.13}$$

and notice that $|g_{\{A,B\}}|_\infty = |g_{\{B,A\}}|_\infty$. To get a better feel for this quantity, we note that $|g_{\{A,B\}}|_\infty = \varepsilon$ means that ν is ϵ-close to being a product distribution in a strong uniform sense. Specifically, let $\widehat{\nu}$ be the product distribution defined by $\widehat{\nu}(\sigma) \equiv \nu(\sigma_B)\cdot\nu(\sigma_A)\cdot\nu^\sigma_{\overline{(A\cup B)}}(\sigma)$. Observe that we can always write $\nu(\sigma) = \nu(\sigma_B)\cdot\nu^\sigma_{\overline{B}}(\sigma_A)\cdot\nu^\sigma_{\overline{(A\cup B)}}(\sigma)$. Thus, since $(1-|g_{\{A,B\}}|_\infty)\nu(\sigma_A) \le \nu^\sigma_{\overline{B}}(\sigma_A) \le (1-|g_{\{A,B\}}|_\infty)^{-1}\nu(\sigma_A)$ for every σ by definition of $|g_{\{A,B\}}|_\infty$, we conclude that $(1-|g_{\{A,B\}}|_\infty)\widehat{\nu}(\sigma) \le \nu(\sigma) \le (1-|g_{\{A,B\}}|_\infty)^{-1}\widehat{\nu}(\sigma)$ for every σ.

Once again, we use lack of correlations (in this case, $|g|_\infty$ close to 0) to establish well-decomposition of entropy (i.e., $\mathcal{E}$ close to 1):

Theorem A.7 *For every two subsets A and B that cover Ψ,*

$$\mathcal{E}_{(A,B)} \;\ge\; (1-|g_{\{\overline{A},\overline{B}\}}|_\infty)^2.$$

Bounds as in Theorem A.7 were proven independently in [Ces01] and [DPP02]. Our bound is slightly sharper, and again the main difference is that it is non-trivial for any value of $|g_{\{\overline{A},\overline{B}\}}|_\infty$, rather than only for small enough values. In addition, although some of the ideas we use in the proof are taken from [Ces01], our proof is much cleaner than both of the previous ones.

Proof of Theorem A.7: We first prove that, if the configurations in $\overline{A}$ and $\overline{B}$ are independent of each other (i.e., $\nu = \widehat{\nu}$, where $\widehat{\nu}$ is defined w.r.t. $\overline{A}$ and $\overline{B}$), then $\mathrm{Ent}(f) \le$

$\nu[\mathrm{Ent}_A(f)] + \nu[\mathrm{Ent}_B(f)]$ for every non-negative f, i.e., $\mathcal{E} = 1$. For this, we use the following two basic properties of entropy. First, by definition, $\mathrm{Ent}(f) = \nu(f \log \frac{f}{\nu(f)})$ and $\nu[\mathrm{Ent}_A(f)] = \nu(f \log \frac{f}{\nu_A(f)})$. Second, by the variational characterization of entropy we have $\nu_A(f \log \frac{g}{\nu_A(g)}) \leq \mathrm{Ent}_A(f)$ for all non-negative functions f and g. We thus have:

$$\begin{aligned} \mathrm{Ent}(f) &= \nu\left[f \log \frac{f}{\nu_A(f)}\right] + \nu\left[f \log \frac{\nu_A(f)}{\nu_B(\nu_A(f))}\right] + \nu\left[f \log \frac{\nu_B(\nu_A(f))}{\nu(f)}\right] \\ &\leq \nu\left[\mathrm{Ent}_A(f)\right] + \nu\left[\mathrm{Ent}_B(f)\right] + \nu\left[f \log \frac{\nu_B(\nu_A(f))}{\nu(f)}\right]. \end{aligned} \tag{A.14}$$

We now notice that the last term on the r.h.s. is 0 when $\nu = \widehat{\nu}$ because then $\nu_B^\sigma(\nu_A(f)) = \nu(\nu_A(f)) = \nu(f)$ for every σ in the support of ν. This completes the proof for the case $\nu = \widehat{\nu}$.

We now go on to consider general ν. The main observation we use is that if ν_1 and ν_2 are two distributions that differ by a factor of at most C, i.e., $\nu_2(\sigma) \leq C\nu_1(\sigma)$ for every σ, then $\mathrm{Ent}_2(f) \leq C\mathrm{Ent}_1(f)$ for every non-negative function f, where Ent_1 and Ent_2 stand for the entropy of f w.r.t. ν_1 and ν_2 respectively. This observation was made previously in, e.g., [Mar98], and follows from the fact that $\mathrm{Ent}(f) = \nu[f\log(f) - f\log(\nu(f)) - f + \nu(f)]$ and that $a\log a - a\log b - a + b \geq 0$ for all non-negative real numbers a, b. (The second fact follows from straightforward calculus.)

We can now conclude the proof as follows. Let $\widehat{\mathrm{Ent}}(f)$ stand for the entropy of f w.r.t. $\widehat{\nu}$. We then write:

$$\begin{aligned} \mathrm{Ent}(f) &\leq \widehat{\mathrm{Ent}}(f) \cdot (1 - |g_{\{A,B\}}|_\infty)^{-1} \\ &\leq (\widehat{\nu}[\widehat{\mathrm{Ent}}_A(f)] + \widehat{\nu}[\widehat{\mathrm{Ent}}_B(f)]) \cdot (1 - |g_{\{A,B\}}|_\infty)^{-1} \\ &= (\nu[\widehat{\mathrm{Ent}}_A(f)] + \nu[\widehat{\mathrm{Ent}}_B(f)]) \cdot (1 - |g_{\{A,B\}}|_\infty)^{-1} \\ &\leq (\nu[\mathrm{Ent}_A(f)] + \nu[\mathrm{Ent}_B(f)]) \cdot (1 - |g_{\{A,B\}}|_\infty)^{-2}. \end{aligned}$$

In the first inequality, we used the fact that $\nu(\sigma) \leq (1 - |g_{\{A,B\}}|_\infty)^{-1}\widehat{\nu}(\sigma)$ for every σ. In the second, we used the fact that the configurations on $\overline{A}$ and $\overline{B}$ are independent of each other under $\widehat{\nu}$. The equality follows from the fact that $\mathrm{Ent}_A(f)$ and $\mathrm{Ent}_B(f)$ depend only on the configurations on $\overline{A}$ and $\overline{B}$ respectively, and that $\nu(\sigma_{\overline{A}}) = \widehat{\nu}(\sigma_{\overline{A}})$ and $\nu(\sigma_{\overline{B}}) = \widehat{\nu}(\sigma_{\overline{B}})$ for every σ. Finally, the last inequality follows from the fact that for every σ in the support of ν, $\widehat{\nu}_A^\sigma(\xi) \leq (1 - |g_{\{A,B\}}|_\infty)^{-1}\nu_A^\sigma(\xi)$ for all ξ, and similarly when A is replaced by B.

We conclude that $\mathcal{E} \geq (1 - |g_{\{A,B\}}|_\infty)^2$, as required. □

We now go on to establish a correspondence between $\mathcal{E}$ and $\mathcal{N}$. Recall that $\mathcal{N}_{(A,B)}$ is not necessarily symmetric in the two subsets A, B. We first notice that by replacing Var

with Ent in the proof of Theorem A.4(i), we get that for any two subsets A, B that cover Ψ, $\max\left\{\mathcal{N}_{(\overline{A},\overline{B})}, \mathcal{N}_{(\overline{B},\overline{A})}\right\} \leq 1 - \mathcal{E}_{\{A,B\}}$. The bound we get for the more interesting converse direction is more involved. Let $p_A = \min\{p = \nu(\sigma_A) : p > 0\}$ stand for the minimum non-zero probability of a configuration on A.

Theorem A.8 *For any two subsets A, B that cover Ψ,*

$$\mathcal{E}_{\{A,B\}} \geq 1 - p_{\overline{A}}^{-1}\sqrt{\mathcal{N}_{(\overline{A},\overline{B})}}\,.$$

It is an interesting open question whether the dependence on $p_{\overline{A}}^{-1}$ in the above bound can be replaced by a dependence on $\mathcal{N}_{(\overline{B},\overline{A})}$, and in particular, whether $\mathcal{E}_{\{A,B\}}$ can be bounded in terms of $\max\left\{\mathcal{N}_{(\overline{A},\overline{B})}, \mathcal{N}_{(\overline{B},\overline{A})}\right\}$ regardless of the minimum probabilities.

Proof of Theorem A.8: Recall that we need to show that, for every non-negative function f,

$$\nu[\mathrm{Ent}_A(f)] + \nu[\mathrm{Ent}_B(f)] \;\geq\; \left(1 - p_{\overline{A}}^{-1}\sqrt{\mathcal{N}_{(\overline{A},\overline{B})}}\right)\mathrm{Ent}(f).$$

By (A.14) it will be enough to show:

$$\nu\left[\nu_A(f)\log\frac{\nu_A(\nu_B(f))}{\nu(f)}\right] \leq p_{\overline{A}}^{-1}\sqrt{\mathcal{N}_{(\overline{A},\overline{B})}}\,\mathrm{Ent}(f).$$

We use the following claim in order to get this bound.

Claim A.9 *Let μ be a probability measure over a finite space Ω where the probability of any $\sigma \in \Omega$ is either zero or at least p. Then for any two non-negative functions f and g over Ω we have*

$$\mu\left[f\log\frac{g}{\mu(g)}\right] \leq \frac{1}{p}\sqrt{\frac{\mu(f)}{\mu(g)}\cdot\mathrm{Ent}(f)\cdot\mathrm{Ent}(g)}\,,$$

where Ent *is taken w.r.t. to μ.*

Assuming Claim A.9, we conclude that

$$\begin{aligned}
\nu\left[\nu_A(f)\log\frac{\nu_A(\nu_B(f))}{\nu(f)}\right] &\leq \frac{1}{p}\sqrt{\mathrm{Ent}[\nu_A(f)]\cdot\mathrm{Ent}[\nu_A(\nu_B(f))]} \\
&\leq \frac{1}{p}\sqrt{\mathcal{N}_{(\overline{A},\overline{B})}\cdot\mathrm{Ent}[\nu_A(f)]\cdot\mathrm{Ent}[\nu_B(f)]} \\
&\leq \frac{1}{p}\sqrt{\mathcal{N}_{(\overline{A},\overline{B})}}\,\mathrm{Ent}(f),
\end{aligned}$$

where in the second inequality we used the fact that $\nu_B(f)$ does not depend on the configuration on B. We note that, since neither $\nu_A(f)$ nor $\nu_A(\nu_B(f))$ depends on A, the effective

probability space in the above derivation is the marginal over $\overline{A}$, so indeed p can be taken as $p_{\overline{A}}$. This completes the proof of Theorem A.8 assuming Claim A.9.

We go on to prove the claim. Consider two arbitrary non-negative functions f and g. Let χ be the indicator function of the event that $g \geq \mu(g)$. Clearly, $\chi \log \frac{g}{\mu(g)} \geq 0$ while $(1-\chi)\log\frac{g}{\mu(g)} \leq 0$. Also, since $\mu\left[\log\frac{g}{\mu(g)}\right] \leq \log\mu\left[\frac{g}{\mu(g)}\right] = 0$ then $\mu\left[(1-\chi)\log\frac{g}{\mu(g)}\right] \leq -\mu\left[\chi\log\frac{g}{\mu(g)}\right]$. Letting $f_{\max}$ and $f_{\min}$ be the maximum and minimum values of f respectively over configurations with non-zero probability, we get:

$$\begin{aligned}
\mu\left[f\log\frac{g}{\mu(g)}\right] &= \mu\left[\chi f\log\frac{g}{\mu(g)}\right] + \mu\left[(1-\chi)f\log\frac{g}{\nu(g)}\right] \\
&\leq f_{\max}\cdot\mu\left[\chi\log\frac{g}{\mu(g)}\right] + f_{\min}\cdot\mu\left[(1-\chi)\log\frac{g}{\mu(g)}\right] \\
&\leq (f_{\max}-f_{\min})\cdot\mu\left[\chi\log\frac{g}{\mu(g)}\right] \\
&\leq \frac{1}{p}\cdot\|f-\mu(f)\|_1\cdot\mu\left[\chi\left(\frac{g}{\mu(g)}-1\right)\right] \\
&= \frac{1}{2p\cdot\mu(g)}\cdot\|f-\mu(f)\|_1\cdot\|g-\mu(g)\|_1 \\
&\leq \frac{1}{p}\sqrt{\frac{\mu(f)}{\mu(g)}\cdot\mathrm{Ent}(f)\cdot\mathrm{Ent}(g)},
\end{aligned}$$

where we wrote $\|\cdot\|_1$ for the L_1 norm with respect to μ and used the fact that $\|f-\mu(f)\|_1^2 \leq 2\mu(f)\mathrm{Ent}(f)$ for any non-negative function f (see, e.g., [Sal97]). The proof of Claim A.9 is now complete. □

We conclude this section by showing that $\mathcal{N}_{(A,B)}$ is bounded by a certain measure of concentration of $g_{(A,B)}$. (The bound was used in Chapter 5 in order to establish the entropy mixing condition EM for spin systems on trees.) Recall that in Theorem A.6 we bounded $\mathcal{C}_{\{A,B\}}$ in terms of the variance of $g_{(A,B)}$. Here we will need a stronger concentration of $g_{(A,B)}$. Fix an ordered pair (A,B) of non-intersecting subsets, and for σ_A in the support of ν, define

$$\delta_{\sigma_A} = \inf\left\{\delta \;:\; \nu\left[|g_{(A,B)}(\sigma_A,\cdot)-1|>\delta\right] \;\leq\; e^{-2/\delta}\right\}. \tag{A.15}$$

Thus, a small value of δ_{σ_A} means that $g_{(A,B)}$ is tightly concentrated in its second variable around its mean value 1. Now, $\mathcal{N}_{(A,B)}$ is bounded in terms of p_A and the expectation of δ_{σ_A}:

Theorem A.10 *There exists a numerical constant c such that*

$$\mathcal{N}_{(A,B)} \;\leq\; cp_A^{-2}\sum_{\sigma_A}\nu(\sigma_A)\delta_{\sigma_A}\,.$$

Notice that since $\mathcal{N}_{(A,B)}$ is not symmetric, it is important to establish the concentration in the *second* variable of $g_{(A,B)}$. (A concentration in the first variable yields a bound on $\mathcal{N}_{(B,A)}$.)

Proof: Fix an arbitrary non-negative function f that depends only on the configuration on B. We have to show that $\mathrm{Ent}(\nu_{\overline{A}}(f)) \leq \varepsilon\mathrm{Ent}(f)$, where $\varepsilon = cp_A^{-2}\sum_{\sigma_A}\nu(\sigma_A)\delta_{\sigma_A}$. Since $\mathrm{Ent}(f') \leq \mathrm{Var}(f')/\nu(f')$ for every non-negative function f' (see, e.g., [Sal97]), then

$$\mathrm{Ent}[\nu_{\overline{A}}(f)] \;\leq\; \frac{\mathrm{Var}[\nu_{\overline{A}}(f)]}{\nu[\nu_{\overline{A}}(f)]} \;=\; \frac{\sum_{\sigma_A}\nu(\sigma_A)\mathrm{Cov}[g_{(A,B)}(\sigma_A,\cdot)\,,\,f(\cdot)]^2}{\nu(f)},$$

where the equality is by (A.10). Thus, the proof will be completed once we show that

$$\mathrm{Cov}\big[g_{(A,B)}(\sigma_A,\cdot)\,,\,f\big]^2 \leq c\nu(\sigma_A)^{-2}\cdot\delta_{\sigma_A}\cdot\nu(f)\mathrm{Ent}(f) \tag{A.16}$$

for some numerical constant c.

To establish (A.16) we make use of the following technical lemma.

Lemma A.11 *Let $\{\Omega,\mathcal{F},\mu\}$ be a probability space and let f_1 be a mean-zero random variable such that $\|f_1\|_\infty \leq 1$ and $\mu[\,|f_1| > \delta\,] \leq e^{-2/\delta}$ for some $\delta \in (0,1)$. Let f_2 be a probability density w.r.t. μ, i.e. $f_2 \geq 0$ and $\mu(f_2) = 1$. Then there exists a numerical constant $c > 0$, independent of $\mu,\ f_1,\ f_2$ and δ, such that $\mu(f_1f_2)^2 \leq c\,\delta\mathrm{Ent}_\mu(f_2)$.*

Let us defer the proof of the lemma for now and complete the proof of Theorem A.10. We apply the above lemma with $\mu = \nu$ and

$$f_1 = \frac{g_{(A,B)}(\sigma_A,\cdot)-1}{\|g_{(A,B)}(\sigma_A,\cdot)\|_\infty}\,;\quad f_2 = \frac{f}{\nu(f)},$$

to deduce $\mathrm{Cov}\big[g_{(A,B)}(\sigma_A,\cdot)\,,\,f(\cdot)\big]^2 \leq \|g_{(A,B)}(\sigma_A,\cdot)\|_\infty^2 c\delta_{\sigma_A}\nu(f)\mathrm{Ent}(f)$; noting also that $\|g_{(A,B)}(\sigma_A,\cdot)\|_\infty \leq \|g_{\sigma_A}\|_\infty = 1/\nu(\sigma_A)$, this establishes (A.16) and thus completes the proof of the theorem. □

Proof of Lemma A.11: We split our analysis of $\mu(f_1f_2)^2$ into three cases:

(a) $\mathrm{Ent}_\mu(f_2) \geq \frac{1}{\delta}$;

(b) $\delta < \mathrm{Ent}_\mu(f_2) < \frac{1}{\delta}$;

(c) $\mathrm{Ent}_\mu(f_2) \le \delta$.

Case (a). We simply bound

$$\mu(f_1 f_2)^2 \le \|f_1\|_\infty^2 \mu(f_2)^2 \le 1 \le \delta\, \mathrm{Ent}_\mu(f_2)\,.$$

Case (b). We use the *entropy inequality* (see, e.g., [ABCF*00]), which states that for any $t > 0$,

$$\mu(f_1 f_2) \le \frac{1}{t} \log \mu(e^{t f_1}) + \frac{1}{t} \mathrm{Ent}_\mu(f_2)\,. \tag{A.17}$$

We choose the free parameter t in (A.17) equal to $\sqrt{\mathrm{Ent}_\mu(f_2)/\delta}$. Notice that, by construction, $1 < t < \delta^{-1}$. Using the assumption $\mu(|f_1| > \delta) \le e^{-2/\delta}$ together with $\|f_1\|_\infty \le 1$, we get

$$\begin{aligned}\mu(f_1 f_2)^2 &\le \Big[\frac{1}{t}\log\big(e^{t\delta} + e^{t-2/\delta}\big) + \sqrt{\delta \mathrm{Ent}_\mu(f_2)}\,\Big]^2 \\ &\le \Big[c_1\,\delta + \sqrt{\delta \mathrm{Ent}_\mu(f_2)}\,\Big]^2 \le c_2\,\delta \mathrm{Ent}_\mu(f_2)\end{aligned}$$

for suitable numerical constants c_1, c_2.

Case (c). Again we use the entropy inequality with $t = \sqrt{\mathrm{Ent}_\mu(f_2)/\delta} \le 1$, but we now simply bound the Laplace transform $\mu(e^{t f_1})$ by a Taylor expansion (in t) up to second order:

$$\begin{aligned}\frac{1}{t}\log \mu(e^{t f_1}) &\le \frac{1}{t}\log\Big(1 + e\frac{t^2}{2}\mu(f_1^2)\Big) \le e\frac{t}{2}\big[\,\delta^2 + e^{-2/\delta}\,\big] \\ &= \frac{1}{2} e\big[\,\delta^2 + e^{-2/\delta}\,\big]\sqrt{\mathrm{Ent}_\mu(f_2)/\delta},\end{aligned}$$

which by (A.17) implies

$$\mu(f_1 f_2)^2 \le \Big[\,\frac{e}{2\sqrt{\delta}}\,\big(\,\delta^2 + e^{-2/\delta}\,\big) + \sqrt{\delta}\,\Big]^2 \mathrm{Ent}_\mu(f_2) \le c_3\,\delta \mathrm{Ent}_\mu(f_2)$$

for another numerical constant c_3. □

Appendix B

Entropy mixing from the log-Sobolev constant

In this appendix we prove that if $c_{\rm sob}$ of the Glauber dynamics is bounded (i.e., entropy in the dynamical system decays "fast" with time), then entropy decays exponentially fast in space (i.e., the contraction in entropy when projecting from a subset A onto a subset B is exponentially small in the distance between the two). This was claimed without proof for trees in Chapter 5 (the reverse direction in Theorem 5.10), but in fact holds for all bounded-degree graphs. Here we state and prove the general version. Recall from Appendix A.3 the notation $\mathcal{N}_{(A,B)}$ for the contraction in entropy when projecting a function that depends only on the configuration on B onto A, and p_A for the minimum non-zero probability of a configuration on A. Also, throughout $c_{\rm sob}(\mu_\Psi^\eta)$ denotes $c_{\rm sob}(P)$, where P is the Glauber dynamics for μ_Ψ^η.

Theorem B.1 *Consider an arbitrary (permissive) spin system on an arbitrary graph of maximum degree $b+1$. There exists a numerical constant c, and another constant $\vartheta > 0$ that depends only on b, such that for every region Ψ, any boundary condition η, and any two non-intersecting subsets A, B of Ψ,*

$$\mathcal{N}_{(A,B)} \;\leq\; c p_A^{-3} |A| \, \exp[-\vartheta \cdot c_{\rm sob}(\mu_\Psi^\eta) \cdot {\rm dist}(A,B)].$$

Proof: Fix Ψ, η, and A, B. We will use Theorem A.10 which tells us that in order to bound $\mathcal{N}_{(A,B)}$, it is enough to establish strong concentration of $g_{(A,B)}(\sigma_A, \cdot)$ for every σ_A. Recall the definition of this function from Appendix A.2 and the measure of its concentration δ_{σ_A},

defined in (A.15). We will show that for every σ_A,

$$\delta_{\sigma_A} \;\leq\; \frac{3e^2|A|}{\nu(\sigma_A)}\,\exp[-\vartheta\cdot c_{\rm sob}(\mu_\Psi^\eta)\cdot{\rm dist}(A,B)] \tag{B.1}$$

for some constant $\vartheta>0$ that depends only on b. The theorem will then follow by applying Theorem A.10.

The proof of the concentration result (B.1) is accomplished by establishing a contraction for a high enough norm:

Lemma B.2 *There exists a constant $\vartheta>0$, depending only on the degree b, such that for any function f that depends only on the configuration on A,*

$$\|\mu_{\overline{B}}(f)-\mu_\Psi^\eta(f)\|_{q+1} \;\leq\; \frac{3|A|}{q}\,\|f-\mu_\Psi^\eta(f)\|_\infty\,,$$

where $q=\exp[\vartheta\cdot c_{\rm sob}(\mu_\Psi^\eta)\cdot{\rm dist}(A,B)]$ and norms are taken w.r.t. μ_Ψ^η.

We first assume Lemma B.2 and complete the proof of the theorem. Applying Lemma B.2 with $f=g_{\sigma_A}$, and recalling that $g_{(A,B)}(\sigma_A,\cdot)=\mu_{\overline{B}}(g_{\sigma_A})$, we get that

$$\|g_{(A,B)}(\sigma_A,\cdot)-1\|_{q+1} \;\leq\; \frac{3|A|\cdot\|g_{(A,B)}(\sigma_A,\cdot)-1\|_\infty}{q} \;\leq\; \frac{3|A|}{q\cdot\mu_\Psi^\eta(\sigma_A)}\,.$$

Therefore, using Markov's inequality,

$$\mu_\Psi^\eta\left[|g_{(A,B)}(\sigma_A,\cdot)-1|>\frac{3e^2|A|}{q\cdot\mu_\Psi^\eta(\sigma_A)}\right] \;\leq\; e^{-2(q+1)}.$$

In particular, $\delta_{\sigma_A}\leq\frac{3e^2|A|}{q\cdot\mu_\Psi^\eta(\sigma_A)}$, which establishes (B.1) and concludes the proof of Theorem B.1. □

Remark: A similar claim to Lemma B.2 was proved in [SZ95] in the context of $\mathbb{Z}^d$; we reprove it below for completeness. Also, a similar version for the L_2 norm, and with $c_{\rm sob}$ replaced by $c_{\rm gap}$, was used in [KMP01] to prove that bounded $c_{\rm gap}$ implies exponential decay of variance with distance.

Proof of Lemma B.2: The proof has two main ingredients: the first is a bound on the speed at which information propagates in the Glauber dynamics, while the second is a standard relationship between $c_{\rm sob}$ and so-called "hypercontractivity".

In order to use the hypercontractivity bound, we have to consider the continuous time version of the dynamics mentioned in Chapter 2. Recall that P is the transition matrix

associated with the (discrete time) Glauber dynamics for μ_Ψ^η. The transition kernel at time t of the continuous time chain is defined as $\widehat{P}_t = e^{t(P-I)}$, where I is the identity matrix. Now, let $\widehat{P}_t^\Lambda$ stand for the transition kernel of a modified dynamics where the spins of the sites outside Λ are fixed to their values at time zero (the sites inside Λ being updated according to the same rule as in the original dynamics). It is well known (see, e.g., [SZ95]) that there exists a constant k_0 depending only on b such that for any subset A, any function f that depends only A, any t and any subset B at distance at least $k_0 t$ from A,

$$\|\widehat{P}_t f - \widehat{P}_t^{\overline{B}} f\|_\infty \leq 2|A| e^{-t} \|f\|_\infty. \tag{B.2}$$

Equation (B.2) is a manifestation of the fact that it takes at least $\frac{\ell}{k_0}$ time before the spin at a site can become sensitive to the configuration at distance ℓ from it, and is based on similar ideas to those we used to prove Lemma 4.4, which gives a bound on the speed of propagation of information in the *discrete time* process.

The second ingredient we need is a hypercontractivity bound. From Gross's integration lemma (see, e.g., [ABCF*00]), we have $\|\widehat{P}_t f\|_q \leq \|f\|_2$ for any mean-zero function f, any t, and $2 \leq q \leq 1 + e^{c_{\rm sob}(\mu_\Psi^\eta) t}$. Adding to this the fact that $c_{\rm gap}(\mu_\Psi^\eta) \geq c_{\rm sob}(\mu_\Psi^\eta)$, we may write

$$\|\widehat{P}_t f\|_q = \|\widehat{P}_{t/2}(\widehat{P}_{t/2} f)\|_q \leq \|\widehat{P}_{t/2} f\|_2 \leq e^{-c_{\rm gap}(\mu_\Psi^\eta) t/2} \|f\|_2 \leq e^{-c_{\rm sob}(\mu_\Psi^\eta) t/2} \|f\|_2, \tag{B.3}$$

where $q = 1 + e^{c_{\rm sob}(\mu_\Psi^\eta) t/2}$ and we used the fact that $c_{\rm gap}$ bounds the rate of decay of the L_2 norm.

We now conclude the proof of Lemma B.2 as follows. Without loss of generality, consider an arbitrary function f that depends only on the configuration in A, with $\mu_\Psi^\eta(f) = 0$. Let $\ell = {\rm dist}(A, B)$. Then, for $t = \ell/k_0$ and $q = 1 + e^{c_{\rm sob}(\mu_\Psi^\eta) t/2}$, we have

$$\begin{aligned}
\|\mu_{\overline{B}}(f)\|_q &= \|\mu_{\overline{B}}(\widehat{P}_t^{\overline{B}} f)\|_q \\
&\leq \|\widehat{P}_t^{\overline{B}} f\|_q \\
&\leq \|\widehat{P}_t^{\overline{B}} f - \widehat{P}_t f\|_q + \|\widehat{P}_t f\|_q \\
&\leq 2|A| e^{-t} \|f\|_\infty + e^{-c_{\rm sob}(\mu_\Psi^\eta) t/2} \|f\|_2 \\
&\leq 3|A| \|f\|_\infty e^{-\vartheta c_{\rm sob}(\mu_\Psi^\eta) \ell},
\end{aligned}$$

taking the constant $\vartheta = 1/2k_0$ (and using the fact that $c_{\rm sob} \leq 1$). □